

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
17. April 2003 (17.04.2003)

PCT

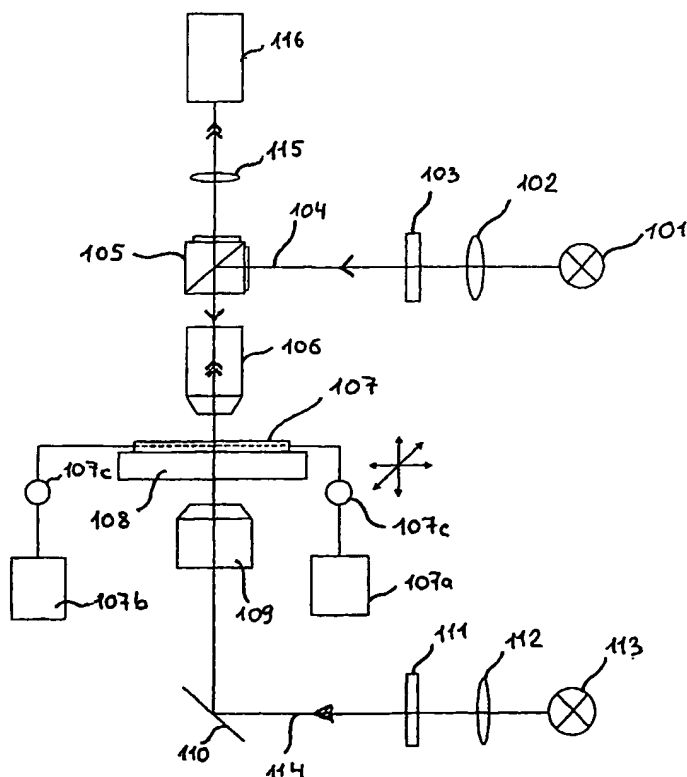
(10) Internationale Veröffentlichungsnummer
WO 03/031947 A2

- (51) Internationale Patentklassifikation: G01N 21/00 (72) Erfinder; und
(21) Internationales Aktenzeichen: PCT/EP02/11098 (75) Erfinder/Anmelder (nur für US): TCHERKASSOV,
Dmitri [RU/DE]; Kahlhorststrasse 36, 23562 Lübeck
(DE).
(22) Internationales Anmeldedatum: 2. Oktober 2002 (02.10.2002) (74) Anwälte: WEBER-QUITZAU, Martin usw.; Uexküll &
Stolberg, Beselerstr. 4, 22607 Hamburg (DE).
(25) Einreichungssprache: Deutsch
(26) Veröffentlichungssprache: Deutsch (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VN, YU, ZA, ZM, ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: DEVICE FOR SEQUENCING NUCLEIC ACID MOLECULES

(54) Bezeichnung: GERÄT ZUR SEQUENZIERUNG VON NUKLEINSÄUREMOLEKÜLEN



(57) Abstract: The invention relates to a device for the automatic determination of nucleic acid sequences. The sequencing reaction occurs by the parallel sequential construction of the strands complementary to individually-fixed single-strand nucleic acid chains. The automatic sequencing carries out said sequential construction and detects electromagnetic radiation from individual marked nucleotides (NT*s) incorporated in the complementary strands. The sequence of the immobilised nucleic acid chain is determined from the order of the incorporated NT*s.

(57) Zusammenfassung: Gegenstand der Erfindung ist ein Gerät zur automatischen Ermittlung von Nukleinsäuresequenzen. Die Sequenzierungsreaktion erfolgt durch den parallelen sequentiellen Aufbau der zu einzelnen fixierten einzelsträngigen Nukleinsäureketten komplementären Stränge. Der Sequenzierautomat führt diesen sequentiellen Aufbau durch und detektiert elektromagnetische Strahlung von einzelnen, markierten in die komplementäre Stränge eingebauten Nukleotiden (NT*s). Aus der Reihenfolge der eingebauten NT*s wird die Sequenz der immobilisierten Nukleinsäureketten bestimmt.

WO 03/031947 A2



(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

Gerät zur Sequenzierung von Nukleinsäuremolekülen

Beschreibung

5 Einführung:

Gegenstand der Erfindung ist ein Gerät zur automatischen Ermittlung von Nukleinsäuresequenzen. Die Sequenzierungsreaktion erfolgt durch den parallelen sequentiellen Aufbau der zu einzelnen fixierten einzelsträngigen Nukleinsäureketten komplementären Stränge. Der Sequenzierautomat führt diesen sequentiellen
10 Aufbau durch und detektiert elektromagnetische Strahlung von einzelnen, markierten in die komplementäre Stränge eingebauten Nukleotiden (NT*s). Aus der Reihenfolge der eingebauten NT*s wird die Sequenz der immobilisierten Nukleinsäureketten bestimmt.

15

1. Abkürzungen und Begriffserläuterungen

DNA - Desoxyribonukleinsäure verschiedenen Ursprungs und unterschiedlicher Länge: (genomische DNA, cDNA, ssDNA, dsDNA)

20

RNA - Ribonukleinsäure (meist mRNA)

Polymerasen - Enzyme, die komplementäre Nukleotide in einen wachsenden DNA- oder RNA-Strang einbauen können (z.B. DNA-Polymerasen, Reverse-
25 Transkriptasen, RNA-Polymerasen)

dNTP – 2'-Desoxy-Nucleosid-Triphosphate als Substrate für DNA-Polymerasen und Reverse-Transkriptasen

30 NT - natürliches Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet.

Die Abkürzung "NT" wird auch bei der Längenangabe einer Nukleinsäuresequenz verwendet, z.B. 1.000 NT. In diesem Fall steht "NT" für Nukleosid-Monophosphate.

- 2 -

Im Text wird bei Abkürzungen die Mehrzahl durch Verwendung des Suffixes "s" gebildet, "NT" steht zum Beispiel für "Nukleotid", "NTs" steht für mehrere Nukleotide.

- 5 NT* - ein mit einem Fluoreszenzfarbstoff und einer zur Termination führenden Gruppe reversibel modifiziertes Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet. NT*s bedeutet: modifizierte Nukleotide

- NSK - steht für eine Nukleinsäurekette (DNA oder RNA). NSKs bedeutet mehrere
10 unterschiedliche oder identische Nukleinsäureketten. NSKs schließen z.B. einzelsträngige oder doppelsträngige Oligo- oder Polynukleotide, genomische DNA, Populationen von cDNAs oder mRNAs ein.

- NSKF - Nukleinsäurekettenfragment, NSKFs - Nukleinsäurekettenfragmente.
15 Fragmente von NSKs (DNA oder RNA), die nach einem Fragmentierungsschritt entstehen. Der Sequenzierautomat kann sowohl für die Analyse von NSKs als auch von NSKFs verwendet werden. Ein wesentlicher Unterschied zwischen NSKs und NSKFs besteht in der Vorbereitung des Materials und in der Analyse der gewonnenen Sequenzen. In der Sequenzierungsreaktion und im Ablauf der
20 Verfahrensschritte bestehen keine große Unterschiede, so dass viele Verfahrensschritt gemeinsam für NSKs und NSKFs beschrieben werden.

- Plane Oberfläche: Oberfläche, die vorzugsweise folgende Merkmale aufweist: 1) Sie erlaubt, mehrere einzelne Moleküle, vorzugsweise mehr als 100, noch besser mehr
25 als 1000, mit dem jeweiligen gegebenen Objektiv-Oberfläche-Abstand bei einer Objektivposition gleichzeitig zu detektieren. 2) Die immobilisierten einzelnen Moleküle befinden sich in derselben Fokusebene, die reproduzierbar eingestellt werden kann.

- 30 Definition der Termination: Als Termination wird in dieser Anmeldung der reversible Stop des Einbaus der modifizierten NT*s bezeichnet. Die modifizierten NT*s tragen

eine reversibel angekoppelte zur Termination führende Gruppe. Diese Gruppe kann von den eingebauten NT*s entfernt werden.

Dieser Begriff ist von dem üblichen Gebrauch des Wortes "Termination" durch Dideoxy-NTP bei einer konventionellen Sequenzierung zu unterscheiden.

5

Genprodukte - mRNA-Transkripte oder von der mRNA abgeleitete Nukleinsäureketten (z.B. einzelsträngige cDNA, von einzelsträngiger cDNA synthetisierte doppelsträngige cDNA, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA). Genprodukte können auch als Gensequenzäquivalente
10 bezeichnet werden.

SNP - single nucleotide polymorphism

PBS – Primerbindungsstelle

15

Objektfeld - ein Teil der Reaktionsoberfläche, der bei einer definierten X,Y-Einstellung des Objektivs durch die Kamera abgebildet werden kann.

Sequenzierungsreaktion – Summe einzelner Verfahrensschritte bis zum Ergebnis:
20 den ermittelten Sequenzen von einzelnen auf der festen Oberfläche immobilisierten NSKs.

DPuMA - Deutsches Patent- und Markenamt

25 2. Stand der Technik

Die am häufigsten verwendete Technik zur Analyse von Nukleinsäuresequenzen ist die Dideoxysequenzierung nach Sanger. Dabei werden markierte Nukleinsäurekettenfragmente nach ihrer Länge in einem Gel aufgetrennt. Ein Beispiel eines solchen Sequenzierautomaten ist in EP 0294524 beschrieben. Dieser
30 Sequenzierautomat kann bis zu 100 Sequenzen gleichzeitig analysieren.

In der vorliegenden Erfindung wird ein Sequenzierapparat präsentiert, der über 100.000 Nukleinsäuresequenzen parallel analysieren kann und somit deutlich größere Sequenziergeschwindigkeit aufweist im Vergleich zu einem „Stand der Technik“ Sequenzierapparat. Dieser Sequenzierautomat ermöglicht sowohl
5 qualitative Analyse von Sequenzen (Sequenzierung im engeren Sinne) als auch eine quantitative Analyse (Auswertung der Anzahl von bestimmten Sequenzen, z.B. bei Genexpressionsanalyse).

Ein solcher Sequenzierautomat kann in vielen Bereichen, z.B. in der Medizin, der Pharmazie und der Biotechnologie eingesetzt werden kann.

10

3. allgemeine Beschreibung

Ein wesentlicher Gegenstand dieser Erfindung ist ein Gerät, ein Sequenzierautomat, zur automatischen parallelen Identifizierung von Nukleinsäuresequenzen. Der erfindungsgemäße Sequenzierautomat kann mehrere
15 hunderttausend einzelner immobilisierter Nukleinsäurenketten parallel sequenzieren. Eine de novo Sequenzierung von Nukleinsäurenketten, eine Analyse von Sequenzvarianten oder auch eine Genexpressionsanalyse sind mit dem beschriebenen Sequenzierautomaten möglich. Dadurch stellt dieser Sequenzierautomat einen Universalautomaten für die Analyse von
20 Nukleinsäuresequenzen dar.

Wesentliche Teile des erfindungsgemäßen Sequenzierautomates sind:

- ein Gehäuse,
- ein optisches System mit Lichtquelle, Filtervorrichtung,
- 25 - Detektionsvorrichtung,
- eine Reaktionsplattform,
- ein Translationssystem (Scantisch),
- und ein Computersystem für die Steuerung einzelner Schritte des Sequenzierungsvorgangs und die Signalanalyse.
- 30 Ein schematisches Beispiel für den Sequenzierautomaten ist in Fig. 1 dargestellt.

- 5 -

Die Sequenzierungsreaktion erfolgt durch den sequentiellen Aufbau der zu den einzelnen fixierten einzelsträngigen Nukleinsäureketten komplementären Stränge. Der Sequenzierautomat führt diesen sequentiellen Aufbau durch und detektiert elektromagnetische Strahlung (Fluoreszenzsignale) von einzelnen, markierten in die
5 komplementären Stränge eingebauten Nukleotiden (NT*s).

Beispiele einer solchen Sequenzierungsreaktion sind in den Anmeldungen Tcherkassov et al. („Verfahren zur Bestimmung der Genexpression“ DPuMA-Aktenzeichen 101 20 798.0-41, „Verfahren zur Analyse von Nukleinsäureketten“ DPuMA-Aktenzeichen 101 20 797.2-41, „Verfahren zur Analyse von
10 Nukleinsäurekettensequenzen und der Genexpression“ DPuMA-Aktenzeichen 101 42 256.3) beschrieben. Die Sequenzierungsreaktion schließt im wesentlichen folgende Schritte ein:

1) Vorbereitung für zyklische Schritte, bestehend aus:

- 15 a) einer Probenvorbereitung, bei der einzelsträngige Nukleinsäureketten (NSKs) mit einer Länge zwischen 20 und 5000 NT, vorzugsweise zwischen 50 und 1000 NT, bereitgestellt und gegebenenfalls mit einer PBS versehen werden, bei längeren Sequenzen wird ein Fragmentierungsschritt durchgeführt, so dass NSKFs entstehen.
- 20 b) einem Fixierungsschritt der vorbereiteten NSK-Probe an die Reaktionsoberfläche in Form von NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen. Dabei werden einzelne NSKs bzw. NSKFs auf der Reaktionsoberfläche in einer solchen Weise fixiert, daß eine enzymatische Reaktion (Synthese des komplementären Stranges) an diesen Molekülen ablaufen kann, s. Beispiel
25 Immobilisation.

2) Nach der Fixierung der NSKs bzw. NSKFs in Form von NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen startet man mit allen auf der Oberfläche immobilisierten Komplexen die zyklischen Schritte. Als Grundlage der
30 Sequenzierung dient die Synthese des komplementären Stranges zu jeder einzelnen fixierten NSK bzw. NSKF. Dabei werden in den neu synthetisierten Strang

- 6 -

markierte NT*s eingebaut. Diese NT*s sind so modifiziert, dass die Polymerase pro Zyklus nur ein einziges markiertes NT* in die wachsende Kette einbauen kann. Diese Modifikation der NT*s ist reversibel, so dass nach der Entfernung dieser Modifikation eine weitere Synthese stattfinden kann. Die Sequenzierungsreaktion
5 verläuft in mehreren Zyklen. Ein Zyklus umfasst im wesentlichen folgende Schritte (zyklische Schritte):

- a) Zugabe einer Lösung mit markierten Nukleotiden (NT*s) und Polymerase zu immobilisierten Nukleinsäureketten;
 - 10 b) Inkubation der immobilisierten Nukleinsäureketten mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
 - c) Waschen,
 - 15 d) Detektion der Signale von einzelnen eingebauten NT*s,
 - e) Entfernung der Fluoreszenz-Markierung und der zur Termination führenden Gruppe von den eingebauten Nukleotiden
 - f) Waschen.
- 20 3) Aus der Reihenfolge der detektierten Signale der eingebauten NT*s wird für jede immobilisierte, und an der Reaktion beteiligte NSK bzw. NSKF die spezifische Sequenz ermittelt.

Ein Beispiel für den allgemeinen Ablauf der Sequenzierungsreaktion ist in Fig. 2
25 dargestellt.

Die im Verfahren verwendbaren NT*s sind reversibel mit einem Farbstoff markiert. Die Auswahlkriterien für diese Farbstoffe sind im Beispiel (Farbstoff) angegeben. Dieser Farbstoff ist an das Nukleotid gekoppelt und kann durch eine chemische
30 oder photochemische Reaktion abgespalten werden. Beispielsweise werden die in den Anmeldungen Tcherkassov et al. („Verfahren zur Bestimmung der

- 7 -

Genexpression" DPuMA-Aktenzeichen 101 20 798.0-41, „Verfahren zur Analyse von Nukleinsäureketten" DPuMA-Aktenzeichen 101 20 797.2-41, „Verfahren zur Analyse von Nukleinsäurekettensequenzen und der Genexpression" DPuMA-Aktenzeichen 101 42 256.3) genannten NT*s eingesetzt. Die genauen Angaben
5 zum Verfahren, zur Synthese und Anwendung von NT*s, einschließlich Polymerasewahl, Reaktionsbedingungen für den NT*-Einbau und Abspaltung sind den oben genannten Quellen dargestellt.

Die Reaktionsbedingungen des Schrittes (b) in einem Zyklus werden so gewählt,
10 daß die Polymerasen an mehr als 50% der an der Sequenzierungsreaktion beteiligten NSKs bzw. NSKFs in einem Zyklus ein markiertes NT* einbauen können, vorzugsweise an mehr als 90%.

Die Anzahl der durchzuführenden Zyklen hängt dabei von der jeweiligen
15 Aufgabenstellung ab, ist theoretisch nicht beschränkt und liegt vorzugsweise zwischen 20 und 5000.

Ein weiterer Gegenstand dieser Erfindung ist eine Reaktionsplattform zur Durchführung von chemischen und biochemischen Reaktionen mit einzelnen
20 Molekülen, insbesondere zur Durchführung von sequentiellen Reaktionen mit einzelnen auf der Oberfläche immobilisierten Nukleinsäureketten. Diese Reaktionsplattform ist vorzugsweise ein Bestandteil des erfindungsgemäßen Sequenzierautomaten.

25 Die Anwendung des Sequenzierautomaten wird an zwei Ausführungsformen verdeutlicht.

In einer Ausführungsform wird der Sequenzierautomat für die Sequenzierung von langen (über 100 kb) Nukleinsäureketten eingesetzt.

Dabei wird aus einer langen Nukleinsäurekette (NSK) eine Population aus relativ kleinen, überlappenden, einzelsträngigen Nukleinsäurekettenfragmenten (NSKFs) erzeugt, diese Fragmente mit einem für den Start der Sequenzierungsreaktion geeigneten Primer versehen, in der Reaktionsplattform fixiert und sequenziert.

- 5 Aus den überlappenden NSKF-Sequenzen kann die ursprüngliche NSK-Sequenz rekonstruiert werden ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press , Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Dabei sucht man in der gesamten Population von
- 10 NSKF-Sequenzen nach Übereinstimmungen/Überlappungen in den Sequenzen von NSKFs. Durch diese Übereinstimmungen/Überlappungen kann man die NSKFs miteinander kombinieren und daraus eine größere zusammenhängende Sequenz rekonstruieren z.B.:

15 CGTCCGTATGATGGTCATTCATG
 CATTCATGGTACGTTAGCTCCTAG
 TCCTAGTAAATCGTACC.

- In der Praxis hat sich bei einer Sequenzierung von unbekannten Sequenzen
- 20 bewährt, eine Länge der sequenzierten Stücke von mehr als 300 bp zu erreichen. Das erlaubt die Sequenzierung von Genomen aus Eukaryonten im Schrotschuss-Verfahren.

- In einer anderen Ausführungsform wird der Sequenzierautomat für die
- 25 Genexpressionsanalyse eingesetzt. Diese Methode basiert auf mehreren Prinzipien:

1. Kurze Nukleotidsequenzen (10-50 NTs) enthalten genügend Informationen zur Identifizierung des korrespondierenden Gens, wenn die Gensequenz selbst bereits in einer Datenbank enthalten ist.
- 30 Eine Sequenz aus beispielsweise 10 NTs kann mehr als 10^6 verschiedene Kombinationen bilden. Das ist z.B. für die meisten Gene im menschlichen Genom,

- 9 -

das nach heutiger Schätzung 32000 Gene enthält, ausreichend. Für Organismen mit weniger Genen kann die Sequenz noch kürzer sein.

2. Der Methode liegt die Sequenzierung einzelner Nukleinsäurekettenmoleküle
5 zugrunde.

3. Es können Nukleinsäureketten-Gemische untersucht werden.

4. Die Sequenzierungsreaktion läuft an vielen Molekülen gleichzeitig ab, wobei die
10 Sequenz jeder einzelnen immobilisierten Nukleinsäurekette analysiert wird.

Es ist bekannt, dass zur Untersuchung der Genexpression mRNAs oder von der mRNA abgeleitete Nukleinsäureketten (z.B. einzelsträngige cDNAs, doppelsträngige cDNAs, von cDNA abgeleitete RNA oder von cDNA amplifizierte
15 DNA) eingesetzt werden können. Unabhängig von der genauen Zusammensetzung werden sie im folgenden als Genprodukte bezeichnet. Auch Teilsequenzen dieser Genprodukte werden im folgenden als Genprodukte bezeichnet.
Diese Genprodukte stellen ein Gemisch aus verschiedenen Nukleinsäureketten dar.

20 Die Genprodukte werden in die einzelsträngige Form überführt, mit einem Primer versehen, auf der Reaktionsoberfläche fixiert und sequenziert.

Die ermittelten Sequenzen der immobilisierten Genprodukte werden zur Bestimmung der Abundanzen untereinander verglichen und durch Vergleich mit
25 Gensequenzen in Datenbanken bestimmten Genen zugeordnet.

3a. detaillierte Beschreibung bevorzugter Ausführungsformen des Sequenzierautomaten:

Apparatur für Detektion

- 5 Zur Detektion der Fluoreszenz von einzelnen Molekülen werden z.B. Nahfeld-Mikroskopie (NFM), Laser-Scanning-Mikroskopie, Totale-Interne-Reflexions-Mikroskopie (TIRM) und Epifluoreszenz-Mikroskopie verwendet. Diese Techniken unterscheiden sich durch ihre physikalischen Prinzipien und die Bauart ihrer optischen Systeme (Science 1999 v.283 1667, Unger et al. BioTechniques 1999
- 10 v.27 S.1008, Ishijama et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Xie et al. Science 1994 v.265 S.361, Nie et al. Science 1994 v.266 S.1018, Betzig et al. Science 1993 v.262 S.1422). Der erfindungsgemäße Sequenzierautomat verwendet den Epifluoreszenzmodus als Mikroskopieprinzip. Dieser Modus wird vorzugsweise verwendet, weil er sich von TIRM, Laser-Scanning
- 15 Microskopie und NFM durch mehrere Vorteile unterscheidet, z.B.:
- 1) die Größe des 2D-Bildes, das z.B. durch eine CCD-Kamera-Aufnahme entsteht und über 1000 Signale von einzelnen Molekülen enthalten kann (z.B. kann man mit einem 100x Objektiv NA 1.4 ein Bild von über 100µm x 100 µm anfertigen)
 - 2) Anregungs- und Fluoreszenzlicht werden durch dasselbe optische System zum

20 Untersuchungsobjekt geleitet. Dadurch wird nur die Objekt-Fläche belichtet, die zur Bildentstehung beiträgt, die benachbarten Regionen werden nicht belichtet.

 - 3) Ein solches System ist kostengünstig im Vergleich zu einem Laser-Scanning-System
- 25 Das Gehäuse, die Translationsvorrichtung (Scantisch), das optische System mit Vergrößerungsvorrichtung (Objektiv), Filtersätzen, dichroischem Spiegel (Farbteiler), Lichtquelle und anderen Hilfsvorrichtungen, wie Lichtquellenkühler, Lichtreflektor, Blenden usw. sind als „Stand der Technik“ Weitfeld-Epifluoreszenzmikroskop kommerziell erhältlich (Firmen: Zeiss, Nikon, Olympus).
- 30 Der schematische Aufbau eines „Stand der Technik“ Epifluoreszenzmikroskop ist in Fig. 3 dargestellt. Ein solches Mikroskop kann in den Sequenzierautomaten

- 11 -

integriert werden. Beispiele sind folgende Mikroskope: Axioskop (Zeiss), Axioplan 2 (Zeiss), Axiovert 100TV, 135TV, 200 (Zeiss), Olympus IX 70 (Olympus), Olympus BX 61 (Olympus), Eclipse TE 300 (Nikon), Eclipse E800 (Nikon). Die genannten Mikroskope dienen für den Fachmann als Beispiele einzelner Bestandteile für den

5 Sequenzierautomaten. Im erfindungsgemäßen Sequenzierautomaten können auch andere gleichwertige funktionelle Einheiten verwendet werden.

Im folgenden werden beispielhaft die wesentlichen Teile der Ausstattung beschrieben.

- 10 Es kann ein aufrechtes oder ein invertiertes Mikroskop eingesetzt werden. Zur Vereinfachung der Darstellung und nicht zur Einschränkung wird im folgenden ein aufrechtes Mikroskop dargestellt (In beiden Fällen wird bevorzugt eine Epifluoreszenzbeleuchtung verwendet; im wesentlichen heißt es, dass das Anregungslicht und das Fluoreszenzlicht durch dasselbe optische System des
- 15 Objektivs geleitet werden)

Es können unterschiedliche Lichtquellen verwendet werden. Dabei kann die Lichtquelle in den Sequenzierautomaten integriert sein oder an ihn über einen Lichtleiter gekoppelt werden.

- 20 Es können Lichtquellen mit kontinuierlichem oder linienförmigem Spektrum verwendet werden. Die spektrale Eigenschaften der Lichtquelle müssen den Anforderungen der Fluoreszenzanregung der Fluoreszenzfarbstoffe entsprechen, s. Beispiel Farbstoffe. Sowohl sichtbares als auch infrarotes Licht kann zur Anregung verwendet werden, wobei eine Lichtquelle sowohl für einen als auch für mehrere
- 25 Farbstoffe verwendet werden kann.

Die Intensität des Anregungslichts bei definierter Wellenlänge liegt zwischen 10 W/cm^2 und 1000.000 W/cm^2 , vorzugsweise zwischen 100 W/cm^2 und 100.000 W/cm^2 auf der beleuchteten Reaktionsoberfläche ($10.000 \text{ W/cm}^2 = 10 \text{ mW} / 100 \mu\text{m}^2$).

Als Lichtquelle dient in einer bevorzugten Ausführungsform eine Lampe. Es können beispielsweise Xe, Hg, Hg/Xe oder „metal halide“ Bogenlampen verwendet werden, z.B. Quecksilberdampf-kurzbogenlampe HBO 50, HBO 100 oder HBO 200. Der Einsatz einer Lampe ist einem Laser vorzuziehen, weil:

- 5 1) das Objektfeld, von dem das 2D-Bild (z.B. 100µm x 100µm) gemacht wird, nahezu gleichmäßig ausgeleuchtet wird,
- 2) Licht mit verschiedenen Wellenlängen erzeugt wird, so dass eine Lampe zur Anregung der Fluoreszenz von mehreren Farbstoffen eingesetzt werden kann.
- 3) Lampen im Vergleich zu Lasern kostengünstiger UV-Licht erzeugen können.

10

In einer anderen bevorzugten Ausführungsform werden ein oder mehrere Laser verwendet (z.B. ein Nd:YAG-Laser, Antares, Coherent, mit doppelten Frequenz, 532nm, zur Anregung von Cy3-Farbstoff und ein Nd:YAG pumped dye Laser, Coherent 700, zur Anregung bei 630nm für Cy5-Farbstoff). Der Vorteil eines Laser

15 besteht in seiner längeren Lebensdauer und einer großen Intensität des Anregungslichtes.

Auch Laser-Dioden können als Lichtquelle verwendet werden.

Die Belichtungszeit liegt vorzugsweise zwischen 0.1 millisekunden (ms) und 20

20 sekunden (s), noch besser zwischen 1 ms und 1s. Sie wird beispielsweise durch einen akustooptischen oder elektrooptischen Modulator oder durch einen „Shutter“ gesteuert, der vom Hauptcomputer kontrolliert wird. Der Shutter kann beispielsweise ein mechanischer Schieber sein.

25 Zur Selektion von Anregungs- und Fluoreszenzlicht und zur Reduktion des Streulichts werden vorzugsweise Filter eingesetzt. Sie sind beispielsweise kommerziell erhältlich (Zeiss, Nikon, Olympus, Leica) und sind an die entsprechenden für die Sequenzierungsreaktion verwendeten Farbstoffe anzupassen. Üblicherweise werden mehrere Filter zu einem Filtersatz

30 zusammengesetzt. Ein solcher Filtersatz besteht üblicherweise aus einem Filter zur Selektion des Anregungslichtes, einem Farbteiler (dichroischem Spiegel) und einem

Filter zur Selektion des Fluoreszenzlichtes. Kommerziell sind sowohl Mono-Band-Filterkombination (für einen Farbstoff, z.B. Cy3 oder Cy5) als auch Multi-Band-Filterkombination (für mehrere Farbstoffe, z.B. Cy3-Cy5-Kombination) erhältlich (z.B. bei Firmen Zeiss, Nikon, Leica, Olympus).

5

Die Filter sind vorzugsweise in einer Halterung befestigt. Diese Halterung ermöglicht einen Austausch zwischen einzelnen Filtern bzw. Filtersätzen. Sowohl ein Filterrevolver als auch ein Filterschieber sind als Stand der Technik bekannt. Der Austausch der Filtersätze erfolgt z.B. automatisch durch einen mit einem Motor angetriebenen Filterrevolver, gesteuert von dem Hauptcomputer.

10

Das Anregungs- und Fluoreszenzlicht wird durch ein Objektiv geleitet. Es werden vorzugsweise PlanNeofluar- und PlanApochromat-Objektive, vorzugsweise Ölimmersionsobjektive, mit einer 40 bis 100 fachen Vergrößerung und mit einer NA von vorzugsweise über 1.2 verwendet, z.B. PlanNeofluar 100x, NA 1.4 (Zeiss), PlanApochromat 100x NA 1.4 (Zeiss), PlanApo100x NA 1.4 Olympus Japan. Vorzugsweise wird Immersionsöl mit niedriger Eigenfluoreszenz verwendet, z.B. Cargille Laboratories, Cedar Grove, NJ, USA. Glycerin oder Wasser können auch als Immersionsmedium mit entsprechenden Immersionsobjektiven verwendet werden.

20

Das Fluoreszenzlicht der eingebauten Nukleotide wird mit einem Objektiv (O) gesammelt und zur Detektionsvorrichtung (D) weitergeleitet. Diese Detektionsvorrichtung stellt vorzugsweise eine gekühlte CCD-Kamera oder intensivierte CCD-Kamera (K) dar. Viele Varianten von Kameras sind kommerziell erhältlich z.B. SenSysTM (von Photometrix), AxioCam (von Zeiss) oder I-PentaMAX (von Roper Scientific, Trenton, NJ, USA).

25

Es werden bevorzugt CCD-chips mit einer hohen Auflösung verwendet. Dies ermöglicht einerseits die Signale von einzelnen Molekülen besser zu identifizieren, bzw. nahe beieinander liegende Signale besser zu differenzieren (s. Beispiel Detektion), andererseits wird bei jeder Aufnahme ein Bild von einer großen Objekt-

30

Fläche und somit eine große Anzahl an Signalen bei genügender Spezifität der Signaleerkennung gleichzeitig aufgenommen.

Moderne Kameras ermöglichen solche Bildaufnahmen und haben CCD-Chips mit einer Auflösung vorzugsweise von mindestens 512x512 Pixel, idealerweise mehr
5 als 1000x1300 Pixel und einer Pixelgröße von ca. 5µm x 5µm.

Es kann sowohl eine Schwarz-Weiß-Kamera (SW-Kamera) als auch eine Farbkamera verwendet werden. Für die SW-Kamera wird Fluoreszenzlicht von gleichen Farbstoffen mit einer Mono-Band-Filterkombination selektiert. Bei einer
10 Farbkamera können Multi-Band-Filterkombinationen eingesetzt werden.

Mit der Kamera wird ein 2D-Bild angefertigt, das Signalintensitäten als Funktion von x,y-Koordinaten wiedergibt. Dieses Bild wird von einem Bildverarbeitungsprogramm analysiert, das sowohl die Signale von eingebauten NT*s vom Hintergrundsignal
15 unterscheiden, als auch nah aneinander liegende Signale differenzieren kann. Ein Beispiel für das Funktionsprinzip eines solchen Programms ist im Beispiel „Detektion“ beschrieben.

Vorzugsweise dient als Translationsvorrichtung ein gesteuerter Scantisch. Solche
20 Tische sind kommerziell erhältlich (Märzhäuser Wetzlar, Zeiss, Leica, Olympus und Nikon). Die Steuerung wird von einem Motor durchgeführt, der durch den Hauptcomputer kontrolliert wird. Diese Tische müssen präzise über mehrere Zyklen dieselben X-Y-Z-Koordinaten einstellen können. Vorzugsweise liegt die Abweichung von einer definierten Position (x-y-z)_i unter 5µm während der gesamten
25 Sequenzierungsreaktion, idealerweise unter 0.1µm.

Der Haupt-Computer (C) ist mit der Detektionsappatur und mit der Reaktionsplattform verbunden und steuert den Ablauf der Sequenzierungsreaktion. Eine möglichst umfassende Automatisierung der Funktionen des
30 Sequenzierautomaten ist angestrebt. Dabei werden vorzugsweise folgende Abläufe bzw. Teile im Sequenzierautomaten automatisiert:

- 15 -

- 1) alle Vorgänge beim Lösungsaustausch an der Reaktionsoberfläche
- 2) alle Vorgänge bei der Detektion
- 3) alle Signalverarbeitungsschritte bis zu Sequenzzusammensetzung

In einer Ausführungsform hat der Hauptcomputer auch Zugang zu genetischen
5 Datenbanken und kann Sequenzzusammensetzung bzw. Sequenzerkennung durchführen.

In Fig. 4a und Fig. 4b sind beispielhafte Ausführungsformen der Detektionsapparatur des erfindungsgemäßen Sequenzierautomaten dargestellt, wobei als Lichtquelle eine
10 Lampe dient.

In Fig. 5 ist eine beispielhafte Anordnung einer Detektionsapparatur mit 2 Lasern dargestellt.

Reaktionsplattform

- 15 Die Reaktionsplattform stellt vorzugsweise eine gesteuerte Durchflussvorrichtung dar. Sie besitzt eine oder mehrere Reaktionsoberflächen und erlaubt einen kontrollierten sequentiellen Austausch von Reaktionslösungen, so dass eine Durchführung von sequenziellen Reaktionen an diesen Oberflächen möglich ist. Im
20 folgenden soll beispielhaft eine Ausführungsform einer solchen Reaktionsplattform dargestellt werden (Fig. 6a).

Es können eine oder mehrere Reaktionsplattformen gleichzeitig verwendet werden. Eine parallele Anordnung von zwei Reaktionsplattformen erlaubt das Scannen der Reaktionsplattform (1) während in der Reaktionsplattform (2) die biochemischen
25 Reaktionen ablaufen. Die Reaktionsplattformen sind am Scantisch befestigt und werden durch diesen bewegt.

In einer bevorzugten Ausführungsform besteht die Reaktionsplattform aus 3 Teilen (Fig. 6a):

- 1) dem austauschbaren Teil, einem Chip (204a) mit einem Mikroflüssigkeitskanal (204b), MFK, der die Reaktionsoberfläche trägt und vorzugsweise für nur eine Sequenzierungsanalyse verwendet wird;
- 2) einem stationären Teil, der Verteilungsvorrichtung (Verteiler) (Fig. 6c), die den Lösungsaustausch im MFK steuert; dabei ist der MFK mit dem Verteiler derart verbunden, dass Lösungen in den MFK automatisch zugeführt und entfernt werden können,
- 3) einem weiteren stationären Teil der Reaktionsplattform, einer Thermostateinheit (Fig. 6c, 220), mit der die Temperatur im MFK geregelt werden kann (Thermoblock).

Ein Beispiel für den Aufbau eines Chips mit dem MFK ist schematisch in Fig. 6b dargestellt. Er besteht aus 2 Platten (222, 223) und 2 Abstandhaltern, so dass ein Kanal (204b) zwischen beiden Platten entsteht. Die Höhe dieses Kanals liegt vorzugsweise zwischen 5 und 200 μm , die Breite zwischen 0.1 und 10 mm und die Länge zwischen 10 und 40 mm. Die dem Objektiv zugewandte Abdeckplatte des MFK besitzt eine für das Anregungs- und Fluoreszenzlicht durchlässige Oberfläche bzw. ein Fenster, vorzugsweise aus Glas. Der Chip selbst kann z.B. aus Glas oder Kunststoff (z.B. PMMA, PVC, Polycarbonate) aufgebaut werden.

In einer anderen Ausführungsform besitzt ein Chip mehrere MFKs (z.B. 2 oder 3 oder 4), wobei der Lösungsaustausch in diesen MFKs unabhängig voneinander gesteuert durch den Verteiler erfolgen kann. Auf diese Weise können unterschiedliche Zyklusschritte parallel in einem Chip ablaufen, so dass die Analysezeit verkürzt wird.

Im folgenden wird als Beispiel ein Chip mit nur einem MFK betrachtet.

Der Austausch von Flüssigkeiten im MFK wird durch den Verteiler (Fig. 6a,c,d,ef) gesteuert. Er besteht in einer Ausführungsform aus einem Bauelement mit integrierten gesteuerten Ventilen, Zuführungsschleuchen und einer oder mehreren Pumpen. Ihre Zahl und genaue Anordnung ist der jeweiligen Ausführungsform anzupassen. Die Flüssigkeitsbeförderung im System erfolgt durch eine oder mehrere

an den Verteiler angeschlossenen durch den Computer gesteuerten Pumpen. Der Verteiler ist mit den Vorratsbehältern der Reaktionslösungen verbunden. Die Ventile regulieren die Zufuhr der Reaktionslösungen. Die Steuerung der Ventile kann beispielsweise mit Motoren, hydraulisch oder elektronisch erfolgen und wird durch
5 den Hauptcomputer kontrolliert.

Je nach Ausführungsform (s. Beispiel Farbstoff, Farbkodierung) werden entweder vier NT*s bzw. zwei NT*s gleichzeitig oder nur ein NT* in die Einbaureaktion zugegeben. Beispielhafte Ausführungsformen sind in Fig. 6d und Fig. 6e
10 angegeben.

In einer Ausführungsform (Fig. 6f) ist ein optischer Detektor zur Kontrolle des Lösungsaustausches integriert. Dieser Detektor ist in den Regelkreis der Reaktionsplattform eingeschaltet und kann beispielsweise durch die Detektion der
15 Veränderungen der durchfließenden Lösung (z.B. optische Dichte, Lichtabsorption oder Fluoreszenz) den Lösungsaustausch kontrollieren.

Bei Bedarf können weitere Modifikationen des Verteilers (zusätzliche Zuführungsschläuche, Pumpen, Ventile usw.) vorgenommen werden, damit auch
20 andere begleitende Schritte der NSK-Sequenzierung automatisch ablaufen können.

Reaktionsoberfläche

Die Reaktionsoberfläche befindet sich vorzugsweise auf der Unterseite der dem
25 Objektiv zugewandten Abdeckplatte des MFK. Die Reaktionsoberfläche ist plan, so dass Signale von vielen einzelnen, auf dieser Oberfläche fixierten Molekülen in der Tiefenschärfe (Fokusebene) des verwendeten Objektivs liegen. Die Anzahl der Signale, die von einem Objektfeld gleichzeitig detektiert werden liegt vorzugsweise über 100, noch bevorzugter über 1000.

Die Reaktionsoberfläche besteht in einer Ausführungsform aus einer festen Phase, z.B. Glas oder Kunststoff (z.B. PMMA) oder Silicon-Derivaten, die für das Anregungs- und Fluoreszenzlicht durchlässig ist. In einer anderen bevorzugten Ausführungsform ist die Reaktionsoberfläche die Oberfläche eines Gels, z.B. eines
5 Polyacrylamidgels. Das Gel liegt auf einer festen Unterlage, z.B. Glas oder Kunststoff, die für das Anregungs- und Fluoreszenzlicht durchlässig ist.

An diese Oberfläche sind die zu sequenzierenden NSKs in Form von NSK-Primer-Komplexen oder NSKF-Primer-Komplexen fixiert, s. Beispiel (Immobilisation). Die
10 Immobilisierungsdichte der NSK-Primer-Komplexe oder NSKF-Primer-Komplexe erlaubt Identifizierung eines einzelnen markierten eingebauten NT-Moleküls auf der Oberfläche. Bevorzugt werden NSK-Primer-Komplexe bzw. NSKF-Primer-Komplexe in einer Dichte immobilisiert, die eine Detektion von mindestens 10 bis 100 Signale pro $100 \mu\text{m}^2$ von einzelnen eingebauten NT*s zulässt bzw. mindestens 50%,
15 idealerweise 90% der identifizierter Fluoreszenzsignale von einzelnen Farbstoffmolekülen stammen, die an die in NSKs eingebauten NT*s gebundenen sind.

Die Reaktionsoberfläche trägt vorzugsweise ein für die Justierung der Bilder
20 geeignetes Muster. Dieses Muster besteht beispielsweise aus Mikroteilchen mit einem Durchmesser von weniger als $1 \mu\text{m}$, die an der Reaktionsoberfläche fixiert sind. Ein Beispiel eines solchen Musters sind auf der Oberfläche fixierte Tusche-Teilchen mit einem Durchmesser von weniger als $1 \mu\text{m}$. Die Dichte der Verteilung dieser Teilchen beträgt vorzugsweise weniger oder gleich 1 Teilchen pro $100 \mu\text{m}^2$.
25 Diese Teilchen dienen erstens zur Einstellung der Fokusebene und zweitens zur Justierung von Bildern (Fluoreszenzbildern) aus verschiedenen Zyklen der Sequenzierungsreaktion (s. Beispiel Detektion).

In einer Ausführungsform können Mikroteilchen Licht absorbieren und werden im
30 Transmissionslicht sichtbar gemacht. In einer anderen Ausführungsform können Mikroteilchen fluoreszieren und werden beispielsweise im Epifluoreszenzmodus

sichtbar gemacht. Unabhängig von der Ausführungsform dürfen diese Teilchen die Reaktion und die Detektion der Fluoreszenzsignale von einzelnen eingebauten NT*s nicht stören.

5 **3b. Ablauf einzelner Schritte im Sequenzierautomaten**

Die Analyse der Sequenzen beinhaltet folgende wesentliche Schritte:

- a) Probenvorbereitung
- b) Immobilisierung von NSKs bzw. NSKFs
- c) Zyklische Schritte
- 10 d) Signalanalyse

Die Probenvorbereitung erfolgt außerhalb des Sequenzierautomaten und ist im Beispiel Probenvorbereitung beschrieben. Die für die Sequenzierungsreaktion vorbereitete NSKs oder NSKFs sind vorzugsweise zwischen 50 und 5000 NT lang
15 und enthalten eine PBS.

Die Schritte b, c und d werden vom Sequenzierautomaten durchgeführt.

Immobilisierung von NSKs bzw. NSKFs:

Ziel dabei ist die NSKs bzw. NSKFs in Form von NSK-Primer-Komplexen bzw.
20 NSKF-Primer-Komplexen auf der Oberfläche zu binden. Dies kann mit verschiedenen Verfahren erfolgen. Einige Beispiele für Fixierung von Komplexen sind im Beispiel (Immobilisierung) angegeben. .

Zyklische Schritte:

25 Die Abfolge der zyklischen Schritte kann sich je nach Ausführungsform unterscheiden. Grundsätzlich werden folgende Schritte in einem Zyklus durchgeführt:

- a) Zugabe einer Reaktionslösung mit markierten Nukleotiden (NT*s)
und Polymerase zu den immobilisierten Nukleinsäureketten,
- 30 b) Inkubation der immobilisierten Nukleinsäureketten mit dieser
Lösung unter Bedingungen, die zur Verlängerung der

- 20 -

- komplementären Stränge um ein NT geeignet sind,
- c) Waschen,
 - d) Detektion der Signale von einzelnen modifizierten, in
die neusynthetisierten Stränge eingebauten NT*-Molekülen,
 - 5 e) Entfernung der Markierung und der zur Termiantion führenden Gruppe
von den eingebauten Nukleotiden,
 - f) Waschen

Zur Vermeidung von unspezifischen Bindung einzelner Komponenten des
10 Reaktionsgemisches kann eine oder mehrere Blockierungslösungen auf die
Oberfläche gebracht werden.

Signalanalyse:

die relative Position einzelner NSKs bzw. NSKFs auf der Reaktionsoberfläche und
15 die Sequenz dieser NSKs bzw. NSKFs werden durch spezifische Zuordnung der in
Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten
Fluoreszenzsignale bestimmt. Diese Signalanalyse und Sequenzrekonstruktion
können parallel zu biochemischen Reaktionen und Detektion oder nach dem
Abschluss der zyklischen Schritte durchgeführt werden. Ein Beispiel für das
20 Funktionsprinzip eines Programms für die Signalanalyse ist im Beispiel Detektion
angegeben.

Der Ablauf der zyklischen Schritt wird vom Hauptcomputer kontrolliert.

25 Unabhängig von dem verwendeten Sequenzierungsverfahren, beispielsweise
(Tcherkassov et al. („Verfahren zur Bestimmung der Genexpression“ DPuMA-
Aktenzeichen 101 20 798.0-41, „Verfahren zur Analyse von Nukleinsäureketten“
DPuMA-Aktenzeichen 101 20 797.2-41, „Verfahren zur Analyse von
Nukleinsäurekettensequenzen und der Genexpression“ DPuMA-Aktenzeichen 101
30 42 256.3), hängt die Wahl der Farbstoffe von dem Filtersystem im

Sequenzierautomaten ab. Einige mögliche Varianten von Farbkodierungen sind im Beispiel (Farbstoffe) dargestellt.

- In einer Ausführungsform können die vier NT*s mit vier unterschiedlichen, aber jeweils spezifischen Farbstoffen markiert werden (z.B. Cy2, Cy3, Cy5, Cy7). In diesem Fall enthält die Reaktionslösung alle vier NT*s . Sie werden im Schritt (b) eingebaut und bilden entsprechend vier verschiedene Signal-Populationen auf der Oberfläche. Zur Detektion der Signale ist in dieser Ausführungsform der Sequenzierautomat mit Filtersätzen ausgestattet, die Selektion der Anregungs- und Fluoreszenzlicht von vier NT*s ermöglichen. Im folgenden wird beispielsweise eine Detektionsvorrichtung eingesetzt, die nur Graustufen-Signal unterscheiden kann, so dass die Farbkodierung der NT*s durch die definierten Filtersatzkombinationen erfolgt.
- Die Signaldetektion in jedem Zyklus erfolgt durch das Abscannen der Oberfläche. Dabei wird die Reaktionsplattform mit der Reaktionsoberfläche durch die Translationsvorrichtung (Scantisch) in X,Y,Z-Achsen bewegt (X,Y-Achse dient dem Positionswechsel, Z-Achse – Einstellung der Fokusebene, s. Beispiel Detektion). Das Abscannen erfolgt so, dass mehrere Felder auf der Oberfläche in einem Zyklus nacheinander untersucht werden, wobei pro Feld mehrere Signale von einzelnen eingebauten NT*s detektiert werden (beispielsweise 5000). Diese Felder stellen vorzugsweise nicht-überlappende Felder dar (Fig.7). In allen Zyklen werden dieselben Felder untersucht. Die Anzahl der Felder, die untersucht werden, hängt von der Gesamtzahl der Sequenzen, die analysiert werden müssen, ab und ist je nach Aufgabestellung unterschiedlich. s. Beispiel Sequenzierung, Genexpresion.

- In einem Zyklus wird jedes Feld mit Anregungslicht für einen bestimmten Farbstoff, selektiv durch den entsprechenden Filtersatz, belichtet. Die Fluoreszenzsignale der eingebauten NT*s werden mit der Detektionsvorrichtung detektiert, so dass je ein 2D-Bild pro Nukleotidart und Objektfeld entsteht. Da die vier NT*s unterschiedliche Markierungen tragen, muss jedes Objektfeld in Kombination mit vier Filtersätzen

nacheinander belichtet wird, so dass vier 2D-Bilder, jeweils mit einem bestimmten Filtersatz, von jedem Objektfeld entstehen. Diese Bilder tragen die Information über die x,y-Verteilung der Signale von eingebauten NT*s. Ein Beispiel eines Programms für die Bildauswertung und Signalerkennung ist im Beispiel Detektion beschrieben.

5

In einer anderen Ausführungsform werden zwei Reaktionsplattformen mit je einem MFK, MFK1 und MFK2, parallel betrieben. Das erlaubt die zeitaufwendigen Teile eines Zyklus parallel durchzuführen: Während in MFK1 die Schritte e-f vom Zyklus n oder die Schritte a-c vom Zyklus n+1 durchgeführt werden, wird im MFK2 der Schritt
10 d durchgeführt, das Abscannen der Reaktionsoberfläche. Dann tauschen MFK1 und MFK2 ihre Positionen und die Reaktionsoberfläche des MFK1 wird abgescannt, während in MFK2 die biochemischen Reaktionen durchgeführt werden.

In einer anderen Ausführungsform werden vier NT*s mit nur zwei verschiedenen
15 Farbstoffen markiert (z.B. Cy3 und Cy5), s. Beispiel (Farbstoffe). Im Zyklus N werden dabei jeweils nur zwei unterschiedlich markierte NT*s gleichzeitig eingesetzt. In dem nächsten Zyklus N+1 werden entsprechend die restlichen zwei unterschiedlich markierte NT*s eingesetzt. In dieser Ausführungsform kann ein Sequenzierapparat mit nur zwei unterschiedlichen Farbfiltern eingesetzt werden.

20 Andere Kombinationen von Farbstoffen, Filtersätzen, sowie Scannen der Oberfläche und der Prozesssteuerung sollten einem Fachmann naheliegend erscheinen.

In einer Ausführungsform wird die Reaktionsoberfläche vor dem ersten Zyklus
25 abgescannt und jedes potentielle Objektfeld wird in Fokus gebracht, wobei die Z-Achsen-Parameter für die Fokuseinstellung jedes Objektfeldes von der Software gespeichert werden. In folgenden Zyklen werden in jedem Detektionsschritt die gespeicherten Z-Achsen-Parameter für jedes Objektfeld verwendet.

In einer anderen Ausführungsform erfolgt eine Fokuseinstellung eines jeden
30 Objektfeldes während des ersten Zyklus, wobei in darauf folgenden Zyklen die gespeicherten Z-Achsen-Parameter für jedes Objektfeld verwendet werden.

- In einer Ausführungsform wird in jedem Zyklus an jedem Objektfeld vor der Detektion der Signale von einzelnen Molekülen eine Kontrolle der Z-Achsen-Einstellung der Reaktionsoberfläche (s. Beispiel Detektion) durchgeführt. Eine
- 5 solche Kontrolle gewährleistet, dass eingebaute NT*s in der Fokusebene des Objektivs liegen und scharf abgebildet werden. Diese Kontrolle wird gleich nach der Einstellung eines neuen Feldes durchgeführt und, falls sich die Oberfläche außerhalb der Fokusebene befindet, wird durch den Z-Antrieb (Beispielsweise des Scantisches, eingebaut in den Mikroskopstativ, oder Piezo-Antrieb des Objektivs),
- 10 die Autofokusfunktion der Software aktiviert und die Oberfläche in den Fokus gebracht. Diese Kontrolle findet auf jedem Objektfeld einmal vor der Aufnahme der Signale von einzelnen Molekülen statt. Mit dieser kontrollierten Z-Position können alle Bilder an diesem Feld in einem Zyklus gemacht werden.
- 15 In einer Ausführungsform wird an jedem Feld in ein Justierungsbild zur Kontrolle der X,Y-Achse-Einstellung der Reaktionsoberfläche gemacht. Ein Justierungsbild kann mit einem im Beispiel Detektion beschriebenen Muster gemacht werden.

- Prinzipien der X,Y,Z-Einstellung eines Objektfeldes sind in einer Ausführungsform
- 20 im Beispiel Detektion dargestellt.

4. Beispiele

4.1 Materialauswahl und Materialvorbereitung:

- Beispiel 4.1.1 Materialauswahl und -vorbereitung bei Sequenzierung langer**
- 25 **NSKs**
- Vorselektionierte DNA-Sequenzen (z.B. in YAC-, PAC-, oder BAC-Vektoren (R. Anand et al. NAR 1989 v.17 S.3425, H. Shizuya et al. PNAS 1992 v.89 S.8794, "Construction of bacterial artificial chromosome libraries using the modified PAC system" in "Current Protocols in Human genetics" 1996 John Wiley & Sons Inc.) klonierte Abschnitte eines
- 30 Genoms) und nicht vorselektionierte DNA (z.B. genomische DNA, cDNA-Gemische) können analysiert werden.

Durch eine Vorselektion ist es möglich, im Vorfeld relevante Informationen, wie z.B. Sequenz-Abschnitte aus einem Genom oder Populationen an Genprodukten, aus der große Menge genetischer Informationen herauszufiltern und damit die Menge der zu analysierenden Sequenzen einzuschränken.

5

Bevorzugt werden gewonnene NSKs ohne Amplifikationsschritte weiter verwendet (z.B. keine PCR und keine Klonierung).

10 Ziel der Materialvorbereitung ist es, gebundene einzelsträngige NSKFs mit einer Länge von vorzugsweise 50-1000 NTs, einer einzelnen Primerbindungsstelle und einem hybridisierten Primer (gebundene NSKF-Primer-Komplexe) zu erhalten. Diese Komplexe können sehr variable Strukturen haben. Zur Verbesserung der Anschaulichkeit folgen nun einige Beispiele, wobei die angeführten Methoden einzeln oder in Kombination eingesetzt werden können.

15

Erzeugung kurzer Nukleinsäurekettenfragmente (50-1000 NTs) (Fragmentierungsschritt); Dieser Schritt wird vorzugsweise außerhalb des Sequenzierautomaten durchgeführt:

20 Wichtig ist, dass die Fragmentierung der NSKs so erfolgt, dass Fragmente erhalten werden, die überlappende Teilsequenzen der Gesamtsequenzen darstellen. Dies wird durch Verfahren erreicht, bei denen unterschiedlich lange Fragmente als Spaltprodukte in zufällmässiger Verteilung entstehen.

Die Erzeugung der Nukleinsäurekettenfragmente (NSKFs) kann durch mehrere
25 Methoden erfolgen, z.B. durch die Fragmentierung des Ausgangsmaterials mit Ultraschall oder durch Endonukleasen ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press), wie z.B. durch unspezifische Endonukleasegemische. Erfindungsgemäß wird die Ultraschall-Fragmentierung bevorzugt. Man kann die Bedingungen so einstellen, dass Fragmente mit einer durchschnittlichen Länge
30 von 100 bp bis 1 kb entstehen. Diese Fragmente können anschließend an ihren Enden durch das Klenow-Fragment (E.coli-Polymerase I) oder durch die T4-DNA-

Polymerase aufgefüllt werden ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press).

Außerdem können aus langen NSKs unter Verwendung randomisierter Primer
5 komplementäre kurze NSKFs synthetisiert werden. Besonders bevorzugt wird diese
Methode bei der Analyse der Gen-Sequenzen. Dabei werden an der mRNA
einzelsträngige DNA-Fragmente mit randomisierten Primern und einer reversen
Transkriptase gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al.
J.Biol.Chem. 1994 v.269 S.31544, Kolls et al. Anal.Biochem. 1993 v.208 S.264,
10 Decraene et al. Biotechniques 1999 v.27 S.962).

Einführung einer Primerbindungsstelle in die NSKFs:

Die Primerbindungsstelle (PBS) ist ein Sequenzabschnitt, der eine selektive Bindung
des Primers an das NSKF ermöglichen soll.

15

In einer Ausführungsform können die Primerbindungsstellen unterschiedlich sein, so
dass mehrere unterschiedliche Primer verwendet werden müssen. In diesem Fall
können bestimmte Sequenzabschnitte der Gesamtsequenz als natürliche PBSs für
spezifische Primer dienen. Diese Ausführungsform ist besonders für die Untersuchung
20 bereits bekannter SNP-Stellen geeignet.

In einer anderen Ausführungsform ist es aus Gründen der Vereinfachung der Analyse
günstig, wenn eine einheitliche Primerbindungsstelle in allen NSKFs vorhanden ist.
Gemäß einer bevorzugten Ausführungsform der Erfindung werden die
Primerbindungsstellen daher in die NSKFs extra eingeführt. Auf diese Weise können
25 Primer mit einheitlicher Struktur für die Reaktion eingesetzt werden.

Im folgenden wird diese Ausführungsform detailliert beschrieben.

Die Zusammensetzung der Primerbindungsstelle ist nicht eingeschränkt. Ihre Länge
beträgt vorzugsweise zwischen 20 und 50 NTs. Die Primerbindungsstelle kann eine
30 funktionelle Gruppe zur Immobilisation des NSKF tragen. Diese funktionelle Gruppe
kann z.B. eine Biotingruppe sein.

Als Beispiel für die Einführung einer einheitlichen Primerbindungsstelle werden im folgenden die Ligation und das Nukleotid-Tailing an DNA-Fragmente beschrieben.

5 a) Ligation:

Dabei wird ein doppelsträngiger Oligonukleotidkomplex mit einer Primerbindungsstelle verwendet. Dieser wird mit kommerziell erhältlichen Ligasen an die DNA-Fragmente ligiert ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press). Es ist wichtig, dass nur eine einzige Primerbindungsstelle an das DNA-
10 Fragment ligiert wird. Das erreicht man z.B. durch eine Modifikation einer Seite des Oligonukleotidkomplexes an beiden Strängen. Die modifizierenden Gruppen am Oligonukleotidkomplex können zur Immobilisation dienen. Die Synthese und die Modifikation eines solchen Oligonukleotidkomplexes kann nach standardisierten Vorschriften durchgeführt werden. Zur Synthese kann z.B. der DNA-Synthesizer 380 A
15 Applied Biosystems verwendet werden. Oligonucleotide mit einer bestimmten Zusammensetzung mit oder ohne Modifikationen sind aber auch als Auftragssynthese kommerziell erhältlich, z.B. von MWG-Biotech GmbH, Germany.

b) Nukleotid-Tailing:

20 Statt der Ligation mit einem Oligonukleotid kann man mit einer terminalen Deoxynucleotidyltransferase mehrere (z.B. zwischen 10 und 20) Nukleosidmonophosphate an das 3'-Ende eines ss-DNA-Fragments anknüpfen ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38), z.B. mehrere Guanosin-Monophosphate ((G)n-
25 Tailing genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (C)n-Primers, verwendet.

Einzelstrang-Vorbereitung:

Für die Sequenzierungsreaktion werden einzelsträngige NSKFs benötigt. Falls das
30 Ausgangsmaterial in doppelsträngiger Form vorliegt, gibt es mehrere Möglichkeiten, aus doppelsträngiger DNA eine einzelsträngige Form zu erzeugen (z.B. Hitze-

Denaturierung oder Alkali-Denaturierung) ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press).

Beispiel 4.1.2 Materialauswahl und -vorbereitung bei der

5 Genexpressionsanalyse:

Genprodukte können von verschiedenen biologischen Objekten stammen, so z.B. von einzelnen Zellen, Zellpopulationen, einem Gewebe oder von kompletten Organismen. Auch biologische Flüssigkeiten wie Blut, Sputum oder Liquor können als Quelle der Genprodukte dienen. Die Methoden zur Gewinnung der Genprodukte aus den
10 verschiedenen biologischen Objekten sind beispielsweise folgenden Literaturquellen zu entnehmen: "Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc..

15 Es kann sowohl die Gesamtheit der isolierten Genprodukte als auch ein durch eine Vorselektion ausgewählter Teil davon in die Sequenzierungsreaktion eingesetzt werden. Durch Vorselektion kann man die Menge der zu analysierenden Genprodukte reduzieren. Die Vorselektion kann beispielsweise durch molekularbiologische Verfahren wie z.B. PCR-Amplifikation, Gel-Auftrennung oder Hybridisierung mit
20 anderen Nukleinsäureketten erfolgen ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.)

Vorzugsweise wird die Gesamtheit der Genprodukte als Ausgangsmaterial gewählt.

Bevorzugt werden Genprodukte ohne Amplifikationsschritte weiter verwendet (z.B.
25 keine PCR und keine Klonierung).

Ziel der Vorbereitung des Materials ist es, aus dem Ausgangsmaterial an die Oberfläche gebundene, extensionsfähige Genprodukt-Primer-Komplexe zu bilden. Wobei pro Genprodukt maximal nur ein Primer binden sollte.

Primerbindungsstelle (PBS):

Jedes Genprodukt hat vorzugsweise nur eine Primerbindungsstelle.

Eine Primerbindungsstelle ist ein Sequenzabschnitt, der eine selektive Bindung des Primers an das Genprodukt ermöglichen soll.

5

Als Primerbindungsstellen können Abschnitte in der Nukleinsäuresequenz dienen, die in den zu analysierenden Sequenzen natürlicherweise vorkommen (z.B. polyA-Strecken in mRNA). Eine Primerbindungsstelle kann auch zusätzlich in das Genprodukt eingeführt werden (Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory , "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.).

Aus Gründen der Vereinfachung der Analyse kann es wichtig sein, dass eine möglichst einheitliche Primerbindungsstelle in allen Genprodukten vorhanden ist.

15 Dann können Primer mit einheitlicher Struktur in die Reaktion eingesetzt werden.

Die Zusammensetzung der Primerbindungsstelle ist nicht eingeschränkt. Ihre Länge beträgt vorzugsweise zwischen 10 und 100 NTs. Die Primerbindungsstelle kann eine funktionelle Gruppe tragen, beispielsweise zur Bindung des Genprodukts an die Oberfläche. Diese funktionelle Gruppe kann z.B. eine Biotin- oder Digoxigenin-Gruppe

20 sein.

Als Beispiel für die Einführung einer Primerbindungsstelle in die Genprodukte wird das Nukleotid-Tailing von antisense cDNA-Fragmenten beschrieben.

25 Als erstes werden einzelsträngige cDNAs von mRNAs synthetisiert. Es resultiert eine Population an cDNA-Molekülen, die eine Kopie der mRNA-Population darstellen, sogenannte antisense-cDNA. (Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory , "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.). Mit einer terminalen

30 Deoxynucleotidyltransferase kann man mehrere (z.B. zwischen 10 und 20) Nukleosid-monophosphate an das 3'-Ende dieser antisense cDNA anknüpfen, z.B. mehrere

Adenosin-Monophosphate ((dA)_n-Tail genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (dT)_n-Primers, verwendet ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38).

5

Beispiel 4.1.3 Primer für die Sequenzierungsreaktion:

Dieser hat die Funktion, den Start an einer einzigen Stelle der NSK bzw. NSKF zu ermöglichen. Er bindet an die Primerbindungsstelle in der NSK (z.B. im Oligonukleotid oder im Genprodukt) oder im NSKF. Die Zusammensetzung und die Länge des Primers sind nicht eingeschränkt. Außer der Startfunktion kann der Primer auch andere Funktionen übernehmen, wie z.B. eine Verbindung zur Reaktionsoberfläche zu schaffen. Primer sollten so an die Länge und Zusammensetzung der Primerbindungsstelle angepaßt werden, dass der Primer den Start der Sequenzierungsreaktion mit der jeweiligen Polymerase ermöglicht.

Bei der Verwendung unterschiedlicher, beispielsweise natürlich in der ursprünglichen Gesamtsequenz vorkommender Primerbindungsstellen, werden die für die jeweilige Primerbindungsstelle sequenzspezifischen Primer verwendet. In diesem Fall wird für die Sequenzierung ein Primergemisch eingesetzt.

Bei einer einheitlichen, beispielsweise durch die Ligation an die NSKFs angekoppelten Primerbindungsstelle wird ein einheitlicher Primer verwendet.

25

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 15 und 30 NTs. Der Primer kann eine Funktionsgruppe tragen, die zur Immobilisierung des NSKF dient, beispielsweise ist eine solche Funktionsgruppe eine Biotingruppe (s. Abschnitt Immobilisierung). Sie soll die Sequenzierung nicht stören. Die Synthese eines solchen Primers kann z.B. mit dem DNA-Synthesizer 380 A Applied Biosystems ausgeführt werden oder aber als

30

Auftragssynthese bei einem kommerziellen Anbieter, z.B. MWG-Biotech GmbH, Germany erstellt werden).

Der Primer kann vor der Hybridisierung an die zu analysierenden NSKs oder NSKFs
5 auf der Oberfläche mit verschiedenen Techniken fixiert oder direkt auf der Oberfläche
synthetisiert werden, beispielsweise nach (McGall et al. US Patent 5412087, Barrett et
al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip
technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena
Oxford University Press, Fodor et al. Science 1991 v.285 S.767, Timofeev et al.
10 Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353,
Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679).

Die Primer werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 bis
100 pro 100 μm^2 , 100 bis 10.000 pro 100 μm^2 oder 10.000 bis 1.000.000 pro 100 μm^2
15 gebunden. Größere Fixierungsdichte wird bevorzugt, wobei keine Notwendigkeit der
optischen Identifizierung eines jeden Primers besteht: größere Primer-dichte
beschleunigt die Hybridisierung von zu analysierenden NSKs oder NSKFs.

Der Primer oder das Primergemisch wird mit NSKFs unter Hybridi-
20 sierungsbedingungen inkubiert, die ihn selektiv an die Primerbindungsstelle der NSKs
oder der NSKFs binden lassen. Diese Primer-Hybridisierung (Annealing) kann vor (1),
während (2) oder nach (3) der Bindung der NSKs bzw. NSKFs an die Oberfläche
erfolgen. Die Optimierung der Hybridisierungsbedingungen hängt von der genauen
Struktur der Primerbindungsstelle und des Primers ab und läßt sich nach Rychlik et al.
25 NAR 1990 v.18 S.6409 berechnen. Im folgenden werden diese
Hybridisierungsbedingungen als standardisierte Hybridisierungsbedingungen
bezeichnet.

Falls eine für alle NSKs bzw. NSKFs gemeinsame Primerbindungsstelle mit bekannter
30 Struktur beispielsweise durch Ligation eingeführt wird, können Primer mit einheitlicher
Struktur eingesetzt werden. Die Primerbindungsstelle kann an ihrem 3'-Ende eine

funktionelle Gruppe tragen, die z.B. zur Immobilisation dient. Beispielsweise ist diese Gruppe eine Biotin-Gruppe. Der Primer hat eine zur Primerbindungsstelle komplementäre Struktur.

- 5 Bindung von Primern an die Oberfläche der MFK erfolgt im Vorfeld zu Experimenten und ist vorzugsweise kein Bestandteil des Verfahrens. Chips mit den an der Oberfläche der MFK gebundenen Primern können längere Zeit gelagert werden.

Beispiel 4.1.4 Immobilisation:

- 10 Fixierung von NSK-Primer-Komplexe bzw. NSKF-Primer-Komplexe an die Oberfläche (Bindung bzw. Immobilisierung von NSKs oder NSKFs).

- Ziel der Fixierung (Immobilisierung) ist es, NSK-Primer-Komplexe bzw. NSKF-Primer-Komplexe auf einer geeigneten planen Oberfläche in einer Art und Weise zu fixieren,
15 dass eine zyklische enzymatische Sequenzierungsreaktion ablaufen kann. Dies kann beispielsweise durch Bindung des Primers (s.o.) oder der NSKs bzw. NSKFs an die Oberfläche erfolgen.

- Die Reihenfolge der Schritte bei der Fixierung von NSK-Primer-Komplexen bzw.
20 NSKF-Primer-Komplexen kann variabel sein:

- 1) Die Komplexe können zunächst in einer Lösung durch Hybridisierung (Annealing) gebildet und anschließend an die Oberfläche gebunden werden.
- 2) Primer können zunächst auf einer Oberfläche gebunden werden und NSKs bzw. NSKFs anschließend an die gebundenen Primer hybridisiert werden, wobei
25 z.B. NSKF-Primer-Komplexe entstehen (NSKFs indirekt an die Oberfläche gebunden)
- 3) Die NSKs bzw. NSKFs können zunächst an die Oberfläche gebunden werden (z.B. NSKFs direkt an die Oberfläche gebunden) und im anschließenden Schritt die Primer an die gebundenen NSKs bzw. NSKFs hybridisiert werden,
30 wobei NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexe entstehen.

Die Immobilisierung der NSKs bzw. NSKFs an die Oberfläche kann daher durch direkte oder indirekte Bindung erfolgen.

In einer bevorzugten Ausführungsform ist die Reaktionsoberfläche ein Bestandteil des MFK, wobei das Material der Oberfläche für die elektromagnetische Strahlung (Anregungs- und Fluoreszenzlicht) durchlässig ist. Dieses Material ist weiterhin enzymatischen Reaktionen gegenüber inert und verursacht keine Störungen der Detektion. Glas oder Kunststoff (z.B. PMMA), oder beliebiges anderes Material, das diesen funktionellen Anforderungen genügt, kann verwendet werden. Vorzugsweise ist die Reaktionsoberfläche nicht verformbar, denn sonst ist mit einer Verzerrung der Signale bei der wiederholten Detektion zu rechnen.

Falls eine gelartige feste Phase (Oberfläche eines Gels) verwendet wird, so kann dieses Gel z.B. ein Agarose- oder Polyacrylamidgel sein. Das Gel ist vorzugsweise für Moleküle mit einer Molekularmasse unter 5000 Da frei passierbar (beispielsweise kann ein 1 bis 2% Agarose-Gel oder 10 bis 15% Polyacrylamid Gel verwendet werden). Eine solche Geloberfläche hat gegenüber festen Reaktionsoberflächen den Vorteil, dass es zu einer wesentlich geringeren unspezifischen Bindung von NT*s an die Oberfläche kommt. Durch die Bindung der NSKF-Primer-Komplexe auf der Oberfläche ist die Detektion der Fluoreszenzsignale von eingebauten NT*s möglich. Die Signale von freien NT*s werden nicht detektiert, weil sie nicht an das Material des Gels binden und somit nicht immobilisiert werden. Das Gel ist vorzugsweise auf einer festen Unterlage befestigt. Diese feste Unterlage kann Glas oder Kunststoff (z.B. PMMA) sein.

Die Dicke des Gels beträgt vorzugsweise nicht mehr als 0,1 mm. Die Geldicke ist vorzugsweise größer als die einfache Tiefenschärfe des Objektivs sein, damit unspezifisch an die feste Unterlage gebundene NT*s nicht in die Fokusebene gelangen und damit detektiert werden. Wenn die Tiefenschärfe z.B. 0,3 µm beträgt, so liegt die Geldicke vorzugsweise zwischen 1 µm und 100 µm. Die Oberfläche kann als eine kontinuierliche Oberfläche oder als diskontinuierliche, aus einzelnen kleinen Bestandteilen (z.B. Agarose-Kügelchen) zusammengesetzte Oberfläche hergestellt

werden. Die Reaktionsoberfläche muß groß genug sein, um die notwendige Anzahl der Komplexen bei entsprechender Dichte immobilisieren zu können. Die Reaktionsoberfläche sollte vorzugsweise nicht größer als 20 cm² sein.

- 5 Falls die Fixierung der NSKF-Primer-Komplexe auf der Oberfläche über die NSKFs erfolgt, kann dies beispielsweise durch die Bindung der NSKFs an einem der beiden Ketten-Enden erfolgen. Dies kann durch entsprechende kovalente, affine oder andere Bindungen erreicht werden. Es sind viele Beispiele der Immobilisierung von Nukleinsäuren bekannt (McGall et al. US Patent 5412087, Nikiforov et al. US Patent
10 5610287, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Rasmussen et al. Analytical Biochemistry v.198, S.138, Allemand et al. Biophysical Journal 1997, v.73, S.2064, Trabesinger et al. Analytical Chemistry 1999, v.71, S.279, Osborne et al. Analytical Chemistry 2000,
15 v.72, S.3678, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679). Die Fixierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der NSKFs enthaltenden Probe auf der planen Oberfläche erreicht werden. Das Gleiche gilt auch für NSKs.

20

Die NSKs bzw. NSKFs werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 und 100 NSks bzw. NSKFs pro 100 µm², 100 bis 10.000 pro 100 µm², 10.000 bis 1.000.000 pro 100µm² gebunden.

- 25 Die für die Detektion notwendige Dichte von extensionsfähigen NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen beträgt ca. 10 bis 100 pro 100 µm². Sie kann vor, während oder nach der Hybridisierung der Primer an die Genprodukte erreicht werden.
- 30 Beispielhaft werden im folgenden einige Methoden zur Bindung von NSKF-Primer-Komplexen näher dargestellt: In einer Ausführungsform erfolgt die Immobilisierung der

NSKFs über Biotin-Avidin oder Biotin-Streptavidin-Bindung. Dabei wird Avidin oder Streptavidin auf der Oberfläche kovalent gebunden, das 5'-Ende des Primers enthält Biotin. Nach der Hybridisierung der markierten Primer mit den NSKFs (in Lösung) werden diese auf der mit Avidin/Streptavidin beschichteten Oberfläche fixiert. Die
5 Konzentration der mit Biotin markierten Hybridisierungs-Produkte sowie die Zeit der Inkubation dieser Lösung mit der Oberfläche wird so gewählt, dass eine für die Sequenzierung geeignete Dichte bereits in diesem Schritt erreicht wird.

In einer anderen bevorzugten Ausführungsform werden die für die
10 Sequenzierungsreaktion geeigneten Primer vor der Sequenzierungsreaktion auf der Oberfläche mit geeigneten Methoden fixiert (s.o.). Die einzelsträngigen NSKs oder NSKFs mit jeweils einer Primerbindungsstelle pro NSK oder NSKF werden damit unter Hybridisierungsbedingungen inkubiert (Annealing). Dabei binden sie an die fixierten Primer und werden dadurch gebunden (indirekte Bindung), wobei Primer-
15 NSK-Komplexe oder Primer-NSKF-Komplexe entstehen. Die Konzentration der einzelsträngigen NSKs oder NSKFs und die Hybridisierungsbedingungen werden so gewählt, dass man eine für die Sequenzierung geeignete Immobilisationsdichte von 10 bis 100 extensionsfähigen Komplexen pro 100 μm^2 erreicht. Nach der Hybridisierung werden ungebundene NSKFs durch einen Waschschrift entfernt. Bei
20 dieser Ausführungsform wird eine Oberfläche mit einer hohen Primerdichte bevorzugt, z.B. ca. 1.000.000 Primer pro 100 μm^2 oder noch höher, da die gewünschte Dichte an NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen schneller erreicht wird, wobei die NSKs bzw. NSKFs nur an einen Teil der Primer binden.

25 In einer anderen Ausführungsform werden die NSKs bzw. NSKFs an die Oberfläche direkt gebunden (s.o.) und anschließend mit Primern unter Hybridisierungsbedingungen inkubiert. Bei einer Dichte von ca. 10 bis 100 NSKs bzw. NSKFs pro 100 μm^2 wird man versuchen alle verfügbaren NSKs bzw. NSKFs mit einem Primer zu versehen und für die Sequenzierungsreaktion verfügbar zu machen.
30 Dies kann z.B. durch hohe Primerkonzentration (Gesamtkonzentration der Primer), beispielsweise 0.1 bis 10 mmol/l, erreicht werden. Bei einer höheren Dichte der

fixierten NSKs bzw. NSKFs auf der Oberfläche, beispielsweise 10.000 bis 1.000.000 pro $100\mu\text{m}^2$, kann die für die optische Detektion notwendige Dichte der Komplexe während der Primer-Hybridisierung erreicht werden. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer, Primerkonzentration) so zu wählen, dass die Primer nur an einen Teil der immobilisierten NSKs bzw. NSKFs binden.

Falls die Oberfläche einer festen Phase (z.B. Silikon oder Glas) zur Immobilisation verwendet wird, wird vorzugsweise eine Blockierungslösung auf die Oberfläche vor dem Schritt (a) in jedem Zyklus gebracht, die zur Vermeidung einer unspezifischen Adsorption von NTs an der Oberfläche dient.

4.2 Beispiel Reaktionslösungen:

- 15 Lösung A: 50 mM Phosphat Puffer pH 8.5, 10% Glycerin, 5 mM Mg^{2+} , 1mM Mn^{2+} ,
Lösung B, (Reaktionslösung $\text{NT}^*(n)$): Lösung A, Polymerase, markierte $\text{NT}^*(n)$
Lösung C, (Abspaltungslösung): Lösung A, Spaltungsreagenzien
Lösung D, die zu analysierende Probe in Lösung A
20 Lösung E (Waschlösung) ist gleich mit Lösung A
Lösung F, 1mg/ml acetyliertes BSA in Lösung A (eine Blockierungslösung zur Reduktion der unspezifischen Bindung der NT^* s an die festen Oberflächen wie Glas, Silicon usw.)

25 4.3 Beispiel Farbstoffe:

Marker, Fluoreszenzfarbstoff

Jede Base ist mit einem für sie charakteristischen Marker (F) markiert. Der Marker
30 ist ein fluoreszierender Farbstoff. Mehrere Faktoren beeinflussen die Wahl des

Fluoreszenzfarbstoffes. Die Wahl ist nicht eingeschränkt, sofern der Farbstoff folgenden Anforderungen genügt:

- a) Die verwendete Detektionsapparatur muß diesen Marker gebunden an DNA unter milden Bedingungen (vorzugsweise Reaktionsbedingungen) als einziges Molekül identifizieren können. Die Farbstoffe haben vorzugsweise große Photostabilität. Ihre Fluoreszenz wird vorzugsweise von der DNA nicht oder nur unwesentlich gequenchet.
 - b) Der an das NT gebundene Farbstoff darf keine irreversible Störung der enzymatischen Reaktion verursachen.
 - c) mit dem Farbstoff markierte NT*s müssen von der Polymerase in die Nukleinsäurekette eingebaut werden.
 - d) Bei einer Markierung mit verschiedenen Farbstoffen sollen diese Farbstoffe keine beträchtlichen Überlappungen in ihren Emissionsspektren aufweisen.
- Im Rahmen der vorliegenden Erfindung verwendbare Fluoreszenzfarbstoffe sind in "Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes mit Strukturformeln zusammengestellt.
- Erfindungsgemäß werden vorzugsweise folgende Farbstoffklassen als Marker eingesetzt: Cyanin-Farbstoffe und deren Abkömmlinge (z.B. Cy2, Cy3, Cy5, Cy7 Amersham Pharmacia Biotech, Waggoner US-Patent 5.268.486), Rhodamine und deren Abkömmlinge (z.B. TAMRA, TRITC, RG6, R110, ROX, Molecular Probes, s. Handbuch), Xanthene-Derivate (z.B. Alexa 568, Alexa 594, Molecular Probes, Mao et al. US-Patent 6.130.101). Diese Farbstoffe sind kommerziell erhältlich.
- Dabei kann man je nach spektralen Eigenschaften und vorhandener Apparatur entsprechende Farbstoffe auswählen. Die Farbstoffe werden an den Linker z.B. über Thiocyanat- oder Ester-Bindung gekoppelt ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Jameson et al. Methods in Enzymology 1997 v.278 S.363, Waggoner Methods in Enzymology

1995 v.246 S.362), s. auch Anmeldungen Tcherkassov et al. („Verfahren zur Bestimmung der Genexpression“ DPuMA-Aktenzeichen 101 20 798.0-41, „Verfahren zur Analyse von Nukleinsäureketten“ DPuMA-Aktenzeichen 101 20 797.2-41, „Verfahren zur Analyse von Nukleinsäurekettensequenzen und der Genexpression“ DPuMA-Aktenzeichen 101 42 256.3).

Farbiges Kodierungsschema, Anzahl der Farbstoffe (Farbkodierung)

Einen Zyklus kann man durchführen mit:

10

- a) vier verschieden markierten NT*s
- b) zwei verschieden markierten NT*s
- c) einem markierten NT*
- d) zwei verschieden markierten NT*s und zwei unmarkierten NTs,

15

d.h.

a) Man kann alle 4 NTs mit verschiedenen Farbstoffen markieren und alle 4 NT*s gleichzeitig in die Reaktion einsetzen. Dabei erreicht man die Sequenzierung einer Nukleinsäurekette mit einer minimalen Anzahl von Zyklen. Diese Variante der Erfindung stellt allerdings hohe Anforderungen an das Detektionssystem: 4 verschiedene Farbstoffe müssen in jedem Zyklus identifiziert werden.

b) Zur Vereinfachung der Detektion kann eine Markierung mit zwei Farbstoffen gewählt werden. Dabei werden 2 Paare von NT*s gebildet, die jeweils verschieden markiert sind, z.B. A und G tragen die Markierung "X", C und U tragen die Markierung "Y". In die Reaktion in einem Zyklus (n) werden 2 unterschiedlich markierte NT*s gleichzeitig eingesetzt, z.B. C* in Kombination mit A*, und im darauffolgenden Zyklus (n+1) werden dann U* und G* zugegeben.

c) Man kann auch nur einen einzigen Farbstoff zur Markierung aller 4 NT*s verwenden und pro Zyklus nur ein NT* einsetzen.

- d) In einer technisch vereinfachten Ausführungsform werden pro Zyklus zwei unterschiedlich markierte NT*s eingesetzt und zwei unmarkierte NTs (sogen. 2NT*s / 2NTs-Methode). Diese Ausführungsform kann verwendet werden, um Varianten
5 (z.B. Mutationen, oder alternativ gespleißte Gene) einer bereits bekannten Sequenz zu ermitteln.

Andere mögliche Kombinationen sind naheliegend.

10 4.4 Beispiel Detektion:

- 1) Vorbereitung zur Detektion
- 2) Durchführung eines Detektionsschrittes in jedem Zyklus. Das Diagramm in Fig. 8 stellt beispielhaft den Ablauf der Detektion an einem Objektfeld, wobei
15 4NT*s (NT*_{1,2,3,4}) mit verschiedenen Farbstoffen markiert sind und in einer Reaktion in die immobilisierten NSKs eingebaut wurden. Jeder Detektionsschritt läuft als Scannvorgang ab und schließt folgende Operationen ein:
 - 20 a) Einstellung der Position des Objektivs (X,Y-Achse),
 - b) Einstellung der Fokusebene (Z-Achse),
 - c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT* und Zuordnung des Signals zur jeweiligen NSK bzw. NSKF,
 - 25 d) Verschiebung zur nächsten Position auf der Oberfläche.

Die Signale von in die NSKs bzw. NSKFs eingebauten NT*s werden durch das Abscannen der Oberfläche registriert. Dabei wird das Objektiv schrittweise über die Oberfläche bewegt (Fig. 7), so dass von jeder Oberflächenposition (Objektfeld) ein zweidimensionales Bild (2D-Bild) entsteht.

1) Vorbereitung zur Detektion:

Bei der Sequenzierung langer Nukleinsäureketten (z.B. 1Mb langen DNA-Stücks) wird am Anfang festgelegt, wie viele NSKFs zur Rekonstruktion der ursprünglichen
5 Sequenz analysiert werden müssen. Im Fall einer Rekonstruktion nach dem Schrotschuß-Verfahren ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257) spielen folgende Faktoren eine Rolle:

- 10 1) Von jedem NSKF wird bei der Sequenzierung eine Sequenz von ca. 300-500 NTs bestimmt.
- 2) Die Gesamtlänge der zu analysierenden Sequenz ist wichtig.
- 3) Bei der Sequenzierung muß ein bestimmtes Maß an Redundanz erreicht werden, um die Genauigkeit zu steigern und eventuelle Fehler zu korrigieren.
- 15 Insgesamt ist zur Rekonstruktion des größten Teils der ursprünglichen Sequenz die etwa 10- bis 100-fache Menge an Rohsequenzen erforderlich, d.h. bei diesem Beispiel mit einer Mb, werden 10 bis 100 Mb Rohsequenzdaten gebraucht. Bei einer durchschnittlichen Sequenzlänge von 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente.

20

Bei der Analyse der Genexpression wird festgelegt, wie viele Kopien der Genprodukte zur Expressionsanalyse notwendig sind. Mehrere Faktoren spielen dabei eine Rolle. Die genaue Zahl hängt z.B. von der relativen Präsenz der Genprodukte im Ansatz und von der gewünschten Genauigkeit der Analyse ab. Die

- 25 Anzahl der analysierten Genprodukte liegt vorzugsweise zwischen 1000 und 10.000.000. Für stark exprimierte Gene kann die Anzahl der analysierten Genprodukte niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse schwach exprimierter Gene muß sie erhöht werden, z.B. auf 100.000 oder noch weiter.

Es werden beispielsweise 100.000 einzelne Genprodukte gleichzeitig analysiert.

- 30 Dabei werden auch schwach exprimierte Gene (mit z.B. ca.100 mRNA-

Molekülen/Zelle, was ca. 0.02% gesamt-mRNA entspricht) in der Reaktion mit durchschnittlich 20 identifizierten Genprodukten repräsentiert.

Die Ermittlung der Gesamtzahl (N_{OF}) der Objektfelder, die abgescannt werden
5 müssen:

Die Anzahl (N_{NSK}) der zu analysierenden NSKs / NSKFs zusammen mit der durchschnittliche Dichte der an der Sequenzierungsreaktion beteiligten NSKs /NSKFs pro Objektfeld (D) bestimmen die Anzahl (N_{OF}) der Objektfelder, die abgescannt werden müssen. Der Hauptcomputer errechnet diese N_{OF} während des
10 ersten Zyklus.

2) Die Durchführung eines Detektionsschrittes in jedem Zyklus wird am Beispiel der Sequenzierung einer langen Nukleinsäurekette erläutert.

Zur Sequenzierung müssen die X,Y-Positionen der NSKFs auf der Oberfläche
15 bestimmt werden, damit man eine Grundlage für die Zuordnung der Signale hat. Die Kenntnis dieser Positionen erlaubt eine Aussage darüber, ob die Signale einzelner Moleküle von eingebauten NT*s stammen oder von zufällig an die Oberfläche gebundenen NT*s. Diese X,Y-Positionen können mit verschiedenen Methoden identifiziert werden.

20

In einer bevorzugten Ausführungsform werden die X,Y-Positionen immobilisierter NSKFs während der Sequenzierung identifiziert. Dabei wird die Tatsache genutzt, daß die Signale von den in die Nukleinsäurekette eingebauten NT*s immer dieselben Koordinaten haben. Das ist durch die Fixierung der Nukleinsäureketten
25 gewährleistet. Die unspezifisch gebundenen NT*s binden zufällig an verschiedene Stellen der Oberfläche.

Zur Identifizierung der X,Y-Positionen von fixierten NSKFs werden die Signale auf Übereinstimmung ihrer Koordinaten aus mehreren aufeinander folgenden Zyklen
30 überprüft. Das kann z.B. am Anfang der Sequenzierung erfolgen. Die

übereinstimmenden Koordinaten werden als Koordinaten der DNA-Fragmente bewertet und gespeichert.

Das Scan-System muß reproduzierbar über mehrere Zyklen die Oberfläche
5 abscannen können. X,Y und Z-Achsen-Einstellungen an jeder Oberflächenposition können von einem Computer kontrolliert werden. Die Stabilität und Reproduzierbarkeit der Einstellung von Objektivpositionen in jedem Scanvorgang entscheiden über die Qualität der Detektion und somit über die Identifizierung der Signale einzelner Moleküle.

10 a) Einstellung der Position des Objektivs (X,Y-Achse) Die mechanische Instabilität der kommerziell erhältlichen Scantische und die geringe Reproduzierbarkeit der wiederholten Einstellung derselben X,Y-Positionen machen eine präzise Analysen der Signale einzelner Moleküle über mehrere Zyklen schwierig. Es existieren viele
15 Möglichkeiten, eine Übereinstimmung der Koordinaten bei wiederholten Einstellungen zu verbessern bzw. mögliche Abweichungen zu kontrollieren. Als Beispiel wird eine Kontrollmöglichkeit angeführt. Nach einer groben mechanischen Einstellung der Objektivposition wird ein Kontrollbild von einem mit der Oberfläche fest verbundenen Muster aufgenommen. Auch wenn die mechanische Einstellung
20 nicht exakt dieselben Koordinaten aufweist (Abweichungen bis zu mehreren μm sind über mehrere Zyklen durchaus möglich), kann man mittels optischer Kontrolle eine Korrektur vornehmen. Das Kontrollbild vom Muster dient als Koordinatensystem für das Bild mit Signalen von eingebauten NT*s. Eine Voraussetzung für eine solche Korrektur ist, daß keine weiteren Bewegungen der
25 Oberfläche zwischen diesen beiden Aufnahmen gemacht werden. Signale von einzelnen Molekülen werden in Relation zum Muster gesetzt, so daß eine X,Y-Abweichung in der Musterposition gleiche X,Y-Abweichung in der Position der Signale einzelner Moleküle bedeutet. Das Kontrollbild vom Muster kann vor, während oder nach der Detektion einzelner Moleküle aufgenommen werden. Ein
30 solches Kontrollbild muß entsprechend bei jeder Einstellung auf einer neuen Oberflächenposition gemacht werden.

b) Einstellung der Fokusebene (Z-Achse)

- 5 Die Oberfläche ist nicht absolut plan und weist verschiedene Unebenheiten auf. Dadurch verändert sich der Oberfläche-Objektiv-Abstand beim abschnen benachbarter Stellen. Diese Unterschiede im Abstand können dazu führen, daß einzelne Moleküle die Fokusebene verlassen und so der Detektion entgehen. Aus diesem Grund ist es wichtig, daß beim Abschnen der Oberfläche die
- 10 Fokusebene korrekt eingestellt wird, bevor eine Aufnahme der Signale von einzelnen Molekülen an jedem Objektfeld durchgeführt wird. Dies geschieht vorzugsweise durch die Einstellung der Fokusebene auf ein bestimmtes Muster, das mit der Reaktionsoberfläche fest verbunden ist. Dieses Muster kann z.B. durch Teilchen mit einem Durchmesser von ca. 1µm gebildet werden. Diese Teilchen
- 15 können z.B. im Durchlicht-Beleuchtungsmodus visualisiert werden. Anschließend wird auf den Fluoreszenzmodus umgeschaltet und Signale von einzelnen Molekülen detektiert.

- In einer Ausführungsform erfolgt die Visualisierung des Einstellungsmuster durch
- 20 die Beleuchtung von unten. Dazu enthält die Reaktionsplattform eine Öffnung in dem unteren Teil, so dass die Reaktionsoberfläche von unten beleuchtet werden kann, z.B. mit Durchlicht- oder Phasenkontrastbeleuchtung (Fig. 4a).

- In einer anderen Ausführungsform kann das Einstellungsmuster selbst fluoreszieren, so dass bei einer entsprechenden Beleuchtung das
- 25 Einstellungsmuster im Fluoreszenzmodus visualisiert werden kann (Fig. 4b).

Vorzugsweise wird für die Visualisierung des Einstellungsmusters das Licht einer anderen Wellenlänge verwendet, die Detektion der Signale von einzelnen Molekülen nicht stört.

- 30 c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT* und Zuordnung des Signals zur jeweiligen NSK.

Das mit Hilfe des Detektionssystems erzeugte zweidimensionale Bild der Reaktionsoberfläche enthält die Signalinformationen von vielen in die NSKF's eingebauten NT's. Diese müssen vor der weiteren Verarbeitung aus der Gesamtdatenmenge der Bildinformationen mit geeigneten Methoden extrahiert
 5 werden. Die dazu notwendigen Algorithmen zur Skalierung, Transformation und Filterung der Bildinformationen zählen zum Standardrepertoire der digitalen Bildverarbeitung und Mustererkennung (Haberäcker P. "Praxis der Digitalen Bildverarbeitung und Mustererkennung". Hanser-Verlag, München, Wien, 1995; Galbiati L.J. "Machine vision and digital image processing fundamentals". Prentice
 10 Hall, Englewood Cliffs, New Jersey, 1990). Die Signalextraktion erfolgt beispielsweise über ein Grauwertbild, das die Helligkeitsverteilung der Reaktionsoberfläche für den jeweiligen Fluoreszenzkanal abbildet. Wenn bei der Sequenzierungsreaktion mehrere Nukleotide mit unterschiedlichen Fluoreszenz-Farbstoffen verwendet werden, kann zunächst für jedes verwendete
 15 fluoreszenzmarkierte Nukleotid (A,T,C,G oder U) ein separates Grauwert-Bild erzeugt werden. Dafür können prinzipiell 2 Verfahren angewendet werden:

1. Durch Verwendung von geeigneten Filtern (z.B. Zeiss-Filtersätze) wird für jeden Fluoreszenzkanal ein Grauwertbild erzeugt.
- 20 2. Aus einem aufgenommenen Mehrkanal-Farb-Bild werden mit Hilfe eines geeigneten Algorithmus durch ein Bildverarbeitungsprogramm die relevanten Farbkanäle extrahiert und jeweils als Grauwertbild einzeln weiterverarbeitet. Zur Kanalextraktion wird dabei ein für den jeweiligen Kanal spezifischer Farb-Schwellwertalgorithmus eingesetzt. So entstehen zunächst aus einem Mehrkanal-
 25 Farbbild einzelne Grauwertbilder 1 bis N. Diese Bilder definieren sich wie folgt:

$GB_N = (s(x,y))$ einkanaliges Grauwertbild

$N = \{1, \dots, \text{Anzahl der Fluoreszenzkanäle}\}$.

$M = \{0, 1, \dots, 255\}$ Grauwertmenge

$S = (s(x,y))$ Bildmatrix des Grauwertbildes

30 $x = 0, 1, \dots, L-1$ Bildzeilen

$y = 0, 1, \dots, R-1$ Bildspalten

- 44 -

(x,y) Ortskoordinaten eines Bildpunktes
s(x,y)? M Grauwert des Bildpunktes.

Aus dieser Datenmenge wird nun durch ein geeignetes Programm die relevante
5 Bildinformation extrahiert. Ein solches Programm sollte folgende Arbeitsschritte realisieren:

Für GB_1 bis GB_N durchführen:

10 I. Vorverarbeitung des Bildes, so zum Beispiel gegebenenfalls Reduktion des durch die Digitalisierung der Bildinformation entstandenen Bildrauschens, etwa durch Grauwertglättung.

II. Prüfung jedes Bildpunktes (x,y) des Grauwertbildes, ob dieser Punkt im
15 Zusammenhang mit den ihn umgebenden unmittelbaren und weiter entfernten Nachbarbildpunkten die Eigenschaften eines Fluoreszenzpunktes erfüllt. Diese Eigenschaften hängen unter anderem von der verwendeten Detektionsapparatur und der Auflösung des Grauwertbildes ab. Sie können beispielsweise ein typisches Verteilungsmuster von Helligkeits-Intensitätswerten über einer den Bildpunkt
20 umgebenden Matrix darstellen. Die dazu verwendbaren Methoden der Bildsegmentierung reichen von einfachen Schwellwertverfahren bis hin zur Verwendung neuronaler Netze.

Erfüllt ein Bildpunkt (x,y) diese Anforderungen, dann folgt ein Vergleich mit den
25 Koordinaten von in bisher durchgeführten Sequenzierungszyklen identifizierten NSKFs. Bei einer Übereinstimmung erfolgt die Zuordnung des Signals mit dem aus dem jeweiligen Fluoreszenzkanal hervorgehenden Nukleotid zu diesem NSKF. Signale mit nicht übereinstimmenden Koordinaten werden als Hintergrundsignale bewertet und verworfen. Die Analyse der Signale kann parallel zum Scanvorgang
30 erfolgen.

In einer beispielhaften Ausführung wird ein 8-Bit-Grauwertbild mit einer Auflösung von 1317 x 1035 Pixel verwendet. Um die durch die Digitalisierung entstandenen Veränderungen am Bild zu reduzieren, erfolgt zunächst eine Vorverarbeitung des Gesamtbildes: Jedem Bildpunkt wird der Mittelwert der Helligkeiten seiner
5 acht-Nachbarn zugewiesen. Bei der gewählten Auflösung entsteht dadurch ein für einen Fluoreszenzpunkt typisches Muster eines zentralen Bildpunktes mit dem größten Helligkeitswert und Nachbarbildpunkten mit nach allen Seiten hin abfallenden Helligkeiten. Erfüllt ein Bildpunkt diese Kriterien und überschreitet der zentrifugale Helligkeitsabfall einen bestimmten Schwellenwert (zur Exklusion zu
10 schwacher Fluoreszenzpunkte), dann wird dieser zentrale Bildpunkt als Koordinate eines Fluoreszenzpunktes gewertet.

d) Verschiebung des Objektivs zur nächsten Position auf der Oberfläche. Nach der Detektion der Signale einzelner Moleküle wird das Objektiv über einer anderen
15 Position der Oberfläche positioniert.

Insgesamt kann beispielsweise eine Folge von Aufnahmen mit der Kontrolle der X,Y-Position, der Einstellung der Fokusebene und mit der Detektion einzelner Moleküle bei jeder neuen Objektivposition gemacht werden. Diese Schritte werden
20 durch einen Computer gesteuert.

4.5 Beispiel Sequenzanalyse:

Sequenzanalyse mit 4 markierten NT*s wobei in einer bevorzugten Ausführungsform der Erfindung werden alle vier in die Reaktion eingesetzten NT*s
25 mit verschiedenen Fluoreszenzfarbstoffen markiert und gleichzeitig in die Reaktion eingesetzt. Diese Ausführungsform kann beispielsweise für unten angeführte Analysen verwendet werden.

4.5.A Sequenzierung langer Nukleinsäuren

30 Diesem Verfahren liegt die Rekonstruktion der ursprünglichen Sequenzen nach dem Schrotschuß-Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231

ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). (Dieses Prinzip ist insbesondere bei der Analyse neuer, unbekannter Sequenzen geeignet.)

5.

Sequenzierung eines langen DNA-Stücks

Im folgenden soll anhand der Sequenzierung eines 1Mb langen DNA-Stückes schematisch die Sequenzierung langer Nukleinsäureketten dargestellt werden. Der Sequenzierung liegt das Shotgun-Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Das zu analysierende Material wird für die Sequenzierungsreaktion vorbereitet, indem es in Fragmente von vorzugsweise 50 bis 1000 bp Länge zerlegt wird. Jedes Fragment wird
10 anschließend mit einer Primerbindungsstelle und einem Primer versehen. Dieses Gemisch aus verschiedenen DNA-Fragmenten wird nun auf einer planen Oberfläche fixiert. Die nicht gebundenen DNA-Fragmente werden durch einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der gesamten Reaktionsoberfläche durchgeführt. Zur Rekonstruktion einer 1 Mb langen DNA-
15 Sequenz sollten die Sequenzen von NSKFs vorzugsweise länger als 300 NTs sein, durchschnittlich ca. 400 bp. Da pro Zyklus nur jeweils ein markiertes NT* eingebaut wird, sind mindestens 400 Zyklen zur Sequenzierung notwendig.
20

Insgesamt ist zur Rekonstruktion der ursprünglichen Sequenz die etwa 10- bis 100-
25 fache Menge an Rohsequenzen erforderlich, d.h. 10 bis 100 Mb. Bei einer durchschnittlichen Sequenzlänge von ca. 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente, um mehr als 99,995% der Gesamtsequenz abzudecken.

Die ermittelten NSKF-Sequenzen stellen eine Population von überlappenden
30 Teilsequenzen dar, die sich mit kommerziell erhältlichen Programmen zur Gesamtsequenz der NSK zusammenfügen lassen ("Automated DNA sequencing

and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press , Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR. 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

5 Sequenzierung der Genprodukte am Beispiel der cDNA-Sequenzierung

10 In einer bevorzugten Ausführungsform können statt einer Sequenz mehrere Sequenzen in einem Ansatz analysiert werden. Die ursprünglichen Sequenzen können aus den gewonnen Rohdaten z.B. nach dem Schrotschuß-Prinzip rekonstruiert werden.

15 Zunächst werden NSKFs erzeugt. Man kann z.B. mRNA in eine doppelsträngige cDNA überführen und diese cDNA mit Ultraschall fragmentieren. Anschließend werden diese NSKFs mit einer Primerbindungsstelle versehen, denaturiert, immobilisiert und mit einem Primer hybridisiert. Zu beachten ist bei dieser Variante der Probenvorbereitung, daß die cDNA-Moleküle unvollständige mRNA-Sequenzen darstellen können (Method in Enzymology 1999, v.303, S.19 und andere Artikel in diesem Band, "cDNA library protocols" 1997 Humana Press).

20 Eine andere Möglichkeit bei der Generierung einzelsträngiger NSKFs von mRNA besteht in der reversen Transkription der mRNA mit randomisierten Primern. Dabei werden viele relativ kurze antisense DNA-Fragmente gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al. J.Biol.Chem. 25 1994 v.269 S.31544, Kolls et al. Anal.Biochem. 1993 v.208 S.264, Decraene et al. Biotechniques 1999 v.27 S.962). Diese Fragmente können anschließend mit einer Primerbindungsstelle versehen werden (s.o). Weitere Schritte entsprechen oben beschriebenen Vorgängen. Mit dieser Methode können komplette mRNA-Sequenzen (vom 5'- bis zum 3'-Ende) analysiert werden, da 30 die randomisierten Primer über die gesamte Länge der mRNA binden.

- 48 -

Immobilisierte NSKFs werden mit einer der oben angeführten Ausführungsformen der Sequenzierung analysiert. Da mRNA-Sequenzen wesentlich weniger repetitive Sequenzen aufweisen als z.B. genomische DNA, kann die Anzahl der detektierten Signale der eingebauten NT*s von einem
5 NSKF geringer als 300 sein und liegt vorzugsweise zwischen 20 und 1000. Die Anzahl der NSKFs, die analysiert werden müssen, errechnet sich nach denselben Prinzipien wie bei einer Schrotschuß-Rekonstruktion einer langen Sequenz.

10 Aus NSKF-Sequenzen werden nach den Prinzipien des Schrotschuß-Verfahrens die ursprünglichen Gensequenzen rekonstruiert.

Diese Methode erlaubt die gleichzeitige Sequenzierung von vielen mRNAs ohne vorherige Klonierung.

15

4.5. B Genexpressionsanalyse:

Sequenzanalyse mit 4 markierten NT*s

20 Bei einer bevorzugten Ausführungsform der Erfindung werden alle vier in die Reaktion eingesetzten NT*s mit Fluoreszenzfarbstoffen markiert.

Dabei verwendet man eine der oben genannten farbigen Kodierungsschemata. Die Zahl der ermittelten NTs für jede Sequenz aus einem Genprodukt liegt zwischen 5 und 100, idealerweise zwischen 20 und 50.

25

Analyse

Die gewonnenen Daten (kurze Sequenzen) werden mit Hilfe eines Programms mit bekannten Gensequenzen verglichen. Einem solchen Programm kann z.B. ein BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational
30 Biology" 1995 M.S. Waterman Chapman & Hall).

Durch die Wahl der Methode zur Materialvorbereitung wird unter anderem festgelegt, in welchen Abschnitten der Genprodukte die Sequenzen ermittelt werden und zu welchem Strang (sense oder antisense) sie gehören. Z.B. werden bei der Verwendung der polyA-Strecken als Primerbindungsstelle in mRNA Sequenzen aus

5 NTRs (non-translating-regions) bestimmt. Bei der Verwendung der Methode mit antisense-cDNA als Matrize stammen die ermittelten Sequenzen unter anderem aus den proteinkodierenden Bereichen der Genprodukte.

Bei einer bevorzugten einfachen Variante der Erfindung wird die Genexpression nur

10 qualitativ bestimmt. Dabei ist nur die Tatsache der Expression bestimmter Gene von Bedeutung.

Bei einer anderen bevorzugten Ausführungsform ist eine quantitative Bestimmung der Verhältnisse zwischen einzelnen Genprodukten im Ansatz von Interesse. Es ist

15 bekannt, daß die Aktivität eines Gens in einer Zelle durch eine Population identischer mRNA-Moleküle repräsentiert ist. In einer Zelle sind viele Gene gleichzeitig aktiv und werden dabei unterschiedlich stark exprimiert, was zum Vorhandensein vieler verschiedener unterschiedlich stark repräsentierter mRNA-Populationen führt.

20 Im folgenden wird auf die quantitative Analyse der Genexpression näher eingegangen:

Für eine quantitative Analyse der Genexpression werden die Abundanzen einzelner Genprodukte in der Sequenzierungsreaktion bestimmt. Dabei sind die Produkte

25 stark exprimierter Gene in der Sequenzierungsreaktion häufiger vertreten als die schwach exprimierter Gene.

Nach der Zuordnung der Sequenzen zu bestimmten Genen wird der Anteil der ermittelten Sequenzen für jedes einzelne Gen bestimmt. Gene mit starker Expression haben einen höheren Anteil an der Gesamtpopulation der Genprodukte

30 als Gene mit schwacher Expression.

Die Anzahl der analysierten Genprodukte liegt vorzugsweise zwischen 1000 und 10.000.000. Die genaue Anzahl der zu analysierenden Genprodukte hängt von der Aufgabenstellung ab. Für stark exprimierte Gene kann sie niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse schwach exprimierter Gene muß sie erhöht werden, z.B.

5 auf 100.000 oder höher.

Werden beispielsweise 100.000 einzelne Genprodukte gleichzeitig analysiert, sind auch schwach exprimierte Gene, wie z.B. ca.100 mRNA-Moleküle/Zelle (was ca. 0.02% gesamt-mRNA entspricht), in der Reaktion mit durchschnittlich 20 identifizierten Genprodukten repräsentiert.

10

Als interne Kontrolle der Hybridisierung, der Immobilisation und der Sequenzierungsreaktion läßt sich folgende Methode verwenden:

Es können eine oder mehrere Nukleinsäureketten mit bekannten Sequenzen als Kontrolle eingesetzt werden. Die Zusammensetzung dieser Kontrollsequenzen ist

15 nicht eingeschränkt, sofern sie die Identifizierung der Genprodukte nicht stören. Bei der Sequenzanalyse der mRNA-Proben werden RNA-Kontrollproben, bei der Analyse der cDNA-Proben entsprechend DNA-Kontrollproben eingesetzt. Diese Proben werden vorzugsweise bei allen Schritten mitgeführt. Sie können z.B. nach der mRNA-Isolation zugegeben werden. Im allgemeinen werden die Kontrollproben
20 in gleicher Weise zur Sequenzanalyse vorbereitet wie die zu analysierenden Genprodukte.

Die Kontrollsequenzen werden in bekannten, fest eingestellten Konzentrationen zu den zu analysierenden Genprodukten zugegeben. Konzentrationen der
25 Kontrollproben können unterschiedlich sein, vorzugsweise liegen diese Konzentrationen zwischen 0.01% und 10% der Gesamtkonzentration der zu analysierenden Probe (100%). Beträgt die Konzentration der mRNA beispielsweise 10ng/µl, dann liegen die Konzentrationen von Kontrollproben zwischen 1pg/µl und 1ng/µl.

30

Bei der quantitativen Analyse der Genexpression muß auch die allgemeine metabolische Aktivität der Zellen berücksichtigt werden, insbesondere, wenn ein Vergleich der Expression bestimmter Gene bei verschiedenen äußeren Bedingungen angestrebt wird.

- 5 Die Veränderung im Expressionsniveau eines bestimmten Gens kann als Folge der Veränderung in der Transkriptionsrate dieses Gens oder als Folge einer globalen Veränderung der Genexpression in der Zelle auftreten. Zur Beobachtung der metabolischen Zustände in der Zelle kann man die Expression der sogenannten "House-keeping-Gene" analysieren. Beim Mangel an wichtigen Metaboliten ist
10 beispielsweise das allgemeine Expressionsniveau in der Zelle niedrig, so daß auch konstitutiv exprimierte Gene ein niedriges Expressionsniveau haben.

Im Prinzip können alle konstitutiv exprimierten Gene als "House-keeping-Gene" dienen. Als Beispiele seien das Transferrin-Rezeptor-Gen oder das Beta-Aktin-Gen genannt.

- 15 Die Expression dieser House-keeping-Gene dient somit als Bezugsgröße für die Analyse der Expression anderer Gene. Die Sequenzermittlung und Quantifizierung der Expression der House-keeping-Gene ist vorzugsweise ein Bestandteil des Analyse-Programms für die Genexpression.

20 **4.6 Beispiel Polymerasen:**

Bei der Wahl der Polymerase spielt die Art der verwendeten fixierten Nukleinsäure (RNA oder DNA) eine entscheidende Rolle:

- Falls RNA als NSKs oder NSKFs oder Genprodukt (z.B. mRNA) in die
25 Sequenzierungsreaktion eingesetzt wird, können handelsübliche RNA-abhängige DNA-Polymerasen eingesetzt werden, z.B. AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse Transcriptase ohne RNase-Aktivität. Alle Reverse Transcriptasen müssen von RNase-Aktivität weitgehend frei sein ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory).

Falls DNA als NSKs oder NSKFs oder Genprodukt (z.B. cDNA) verwendet wird, eignen sich als Polymerasen prinzipiell alle DNA-abhängigen DNA-Polymerasen ohne 3'-5' Exonuklease-Aktivität (DNA-Replication" 1992 Ed. A.Kornberg, Freeman and company NY), z.B. modifizierte T7-Polymerase vom Typ "Sequenase Version 2" (Amersham Pharmacia Biotech), Klenow Fragment der DNA-Polymerase I ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech), Polymerase Beta verschiedenen Ursprungs (Animal Cell DNA Polymerases" 1983, Fry M., CRC Press Inc., kommerziell erhältlich bei Chimex) thermostabile Polymerasen wie beispielsweise Taq-Polymerase (GibcoBRL), proHA-DNA-Polymerase (Eurogentec).

10

Polymerasen mit 3'-5' Exonuklease-Aktivität können eingesetzt werden (z.B. Klenow-Fragment der E.coli-Polymerase I), sofern Reaktionsbedingungen gewählt werden, die vorhandene 3'-5' Exonuklease-Aktivität unterdrücken, wie z.B. ein niedriger pH-Wert (pH 6.5) beim Klenow-Fragment (Lehman and Richardson, J. Biol. Chem. 1964 v.239 S.233) oder Zugabe von NaF zur Einbaureaktion. Eine andere Möglichkeit besteht in der Verwendung von NTs* mit einer Phosphorothioate-Verbindung (Kunkel et al. PNAS 1981, v.78 S.6734). Dabei werden eingebaute NTs* von der 3'-5' Exonuklease-Aktivität der Polymerase nicht angegriffen. Im folgenden werden all diese Polymerasearten als "Polymerase" bezeichnet.

20

4.7 Beispiel modifizierte Nukleotide:

Für die hoch parallele Sequenzierung mit einzelnen Molekülen (parallele Sequenzanalyse von bis zu 10.000.000 Nukleinsäuremolekülen) ist wichtig, dass jedes eingebaute NT* während der Sequenzierungsreaktion identifiziert wird. Eine Voraussetzung dafür ist, dass nur ein einziges NT* pro Zyklus in die Nukleinsäurekette eingebaut wird.

25

Das wird erreicht durch eine reversible Kopplung einer zur Termination führenden Gruppe. Diese Gruppe kann sowohl an der Base (z.B. 5-Position der Pyrimidine bzw. 7 Position der 7-Deazapurine) als auch an der 3'-Position der Ribose bzw. 2'-Deoxyribose des Nukleotides gekoppelt sein.

30

Falls diese Gruppe an der Base gekoppelt ist, so ist sie eine sterisch anspruchsvolle Gruppe, die durch ihre chemische Struktur die Eigenschaften der mit dieser Gruppe gekoppelten NTs* so verändert, dass diese durch eine Polymerase in einer Extensionsreaktion nicht nacheinander eingebaut werden können. Wenn ein
5 Reaktionsgemisch, das nur modifizierte NTs* enthält, in der Reaktion eingesetzt wird, dann kann die Polymerase nur ein einziges NT* einbauen. Der Einbau eines nächsten NT* wird sterisch gehemmt. Diese NTs* treten somit als Terminatoren der Synthese auf. Nach der Entfernung der sterisch anspruchsvollen Gruppe kann das nächste komplementäre NT* eingebaut werden. Weil diese NTs* kein absolutes Hindernis zur
10 weiteren Synthese darstellen, sondern nur für den Einbau eines weiteren markierten NT*, werden sie als Semiterminatoren bezeichnet.

Allgemeine Struktur des NT* mit sterischer Hinderung:

Ihre gemeinsamen Merkmale sind in Fig. 9a,b,d dargestellt. Diese Struktur ist dadurch
15 charakterisiert, dass an der Base über einen spaltbaren Linker (A-E) eine sterische Gruppe (D) und der Fluoreszenzmarker (F) gebunden sind.

Als Grundlage für die NTs* dienen Deoxynukleosid-Triphosphate mit Adenosin (A), Guanosin(G), Cytidin (C) und Uridin (U) als Nukleosidrest. Anstelle von Guanosin
20 kann Inosin verwendet werden.

Natur der sterisch anspruchsvollen Gruppe

Die Gruppe (D) (Fig. 9a,b,d) stellt ein Hindernis für den Einbau eines weiteren komplementären markierten NT* durch eine Polymerase dar.

25

Biotin, Digoxigenin und Fluoreszenzfarbstoffe wie Fluoreszein, Tetramethylrhodamine und Cy3-Farbstoff stellen Beispiele einer solchen sterisch anspruchsvollen Gruppe dar (Zhu et al. Cytometry 1997, v.28, S.206, Zhu et al. NAR 1994, v.22, S.3418, Gebeyehu et al., NAR 1987, v.15, S.4513, Wiemann et al. Analytical Biochemistry
30 1996, v.234, S.166, Heer et al. BioTechniques 1994 v.16 S.54). Die chemische Struktur dieser Gruppe ist nicht eingeschränkt, sofern sie den Einbau des markierten

- 54 -

NT*, an das sie geknüpft ist, nicht wesentlich stört und keine irreversible Störung der enzymatischen Reaktion verursacht.

Diese Gruppe kann als selbständiger Teil im Linker (6a) auftreten oder mit dem
5 Farbstoff (9b) oder der spaltbaren Gruppe (9d) identisch sein. Durch die Spaltung des Linkers wird diese sterisch anspruchsvolle Gruppe (D) nach der Detektion des Signals entfernt, so dass die Polymerase ein weiteres markiertes NT* einbauen kann. Bei einer Struktur wie in 6d wird die sterische Gruppe durch die Spaltung beseitigt.

10 In einer bevorzugten Ausführungsform übernimmt der Fluoreszenzfarbstoff die Funktion einer solchen sterisch anspruchsvollen Gruppe, so dass ein markiertes Nukleotid eine in Fig. 9b dargestellte Struktur aufweist.

In einer anderen bevorzugten Ausführungsform übernimmt die photolabile spaltbare
15 Gruppe die Funktion einer solchen sterisch anspruchsvollen Gruppe (Fig. 9d).

Linker:

Der Marker (Fluoreszenzfarbstoff) ist an die Base vorzugsweise über einen Abstandhalter unterschiedlicher Länge, einen sog. Linker, gebunden. Beispiele für
20 Linker sind in Fig. 9e,f,g,h,i,j gegeben. Beispiele der Ankoppelung eines Linkers an die Base können aus folgenden Quellen entnommen werden (Hobbs et al. US Patent 5.047.519, Khan et al. US Patent 5.821.356, Klevan et al. US Patent 4.828.979, Hanna M. Method in Enzymology 1996 v.274, S.403, Zhu et al. NAR 1994 v.22 S.3418, Herman et al. Methods in Enzymology 1990 v.184 S.584, J.L.Ruth et al.
25 Molecular Pharmacology 1981 v.20 S.415, L. Ötvös et al. NAR 1987 v.15 S.1763, G.E.Wright et al. Pharmac Ther. 1990 v.47, S.447, „Nucleotide Analogs; Synthesis and Biological Function“ K.H. Scheit 1980, Wiley-Interscience Publication, „Nucleic acid chemistry“ Ed. L.B.Townsend, v.1-4, Wiley-Interscience Publication, „Chemistry of Nucleosides and Nucleotides“ Ed. L.B.Townsend, v.1-3, Plenum Press).

Die gesamte Länge des Linkers kann variieren. Sie entspricht der Anzahl der Kohlenstoff-Atome in den Abschnitten A, C, E (Fig. 9a,b,d) und liegt vorzugsweise zwischen 3 und 20. Optimalerweise beträgt sie zwischen 4 und 10 Atomen. Die chemische Zusammensetzung des Linkers (Abschnitte A,C,E in Fig. 9a,b,d) ist nicht
5 eingeschränkt, sofern sie unter Reaktionsbedingungen stabil bleibt und keine Störung der enzymatischen Reaktion verursacht.

Spaltbare Verbindung, Spaltung:

- 10 Der Linker trägt eine spaltbare Verbindung oder spaltbare Gruppe (Abschnitt (B) in Fig. 9a,b,d). Diese spaltbare Verbindung ermöglicht die Entfernung des Markers und des sterischen Hindernisses am Ende jedes Zyklus. Ihre Wahl ist nicht eingeschränkt, sofern sie unter den Bedingungen der enzymatischen Sequenzierungsreaktion stabil bleibt, keine irreversible Störung der Polymerase verursacht und unter milden
15 Bedingungen abgespalten werden kann. Unter "milden Bedingungen" sind solche Bedingungen zu verstehen, die den Genprodukt-Primer-Komplex nicht zerstören, wobei z.B. der pH-Wert vorzugsweise zwischen 3 und 11 liegt, die Temperatur zwischen 0°C und einem Temperaturwert (x). Dieser Temperaturwert (x) hängt von der T_m des Genprodukt-Primer-Komplexes (T_m ist "melting Point") und wird
20 beispielsweise als T_m (Genprodukt-Primer-Komplex) minus 5°C errechnet (z.B. T_m ist 47°C, dann liegt die maximale Temperatur bei 42°C; unter diesen Bedingungen eignen sich besonders Ester-, Thioester-, Disulfid-Verbindungen und photolabile Verbindungen als spaltbare Verbindungen).
- 25 Vorzugsweise gehört die genannte Verbindung zu chemisch oder enzymatisch spaltbaren oder photolabilen Verbindungen. Als Beispiele von chemisch spaltbaren Gruppen sind Ester-, Thioester- und Disulfid-Verbindungen bevorzugt („Chemistry of protein conjugation and crosslinking" Shan S. Wong 1993 CRC Press Inc., Herman et al. Method in Enzymology 1990 v.184 S.584, Lomant et al. J.Mol.Biol. 1976 v.104
30 243, "Chemistry of carboxylic acid and esters" S.Patai 1969 Interscience Publ.). Beispiele für photolabile Verbindungen können in folgenden Literaturstellen gefunden

werden: "Protective groups in organic synthesis" 1991 John Wiley & Sons, Inc., V. Pillai Synthesis 1980 S.1, V. Pillai Org.Photochem. 1987 v.9 S.225, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“ H. Giegrich, 1996, Konstanz, Dissertation „Neue photolabile Schutzgruppen für die
5 lichtgesteuerte Oligonucleotidsynthese“ S.M.Bühler, 1999, Konstanz).

Die Position der spaltbaren Verbindung/Gruppe im Linker ist vorzugsweise nicht weiter als 10 Atome von der Base entfernt, noch bevorzugter nicht weiter als 3 Atome. Besonders bevorzugt liegt die spaltbare Verbindung oder Gruppe direkt an der Base.

- 10 Der Spaltungs- und Entfernungsschritt ist in jedem Zyklus vorhanden und muß unter milden Bedingungen (s.o.) verlaufen, so dass die Nukleinsäuren nicht beschädigt oder modifiziert werden.

Die Spaltung läuft bevorzugt chemisch (z.B. in milder saurer oder basischer
15 Umgebung für eine Ester-Verbindung oder durch Zugabe eines Reduktionsmittels, z.B. Dithiothreitol oder Mercaptoethanol (Sigma) bei der Spaltung einer Disulfid-Verbindung), oder physikalisch (z.B. durch Beleuchtung der Oberfläche mit Licht einer bestimmten Wellenlänge für die Spaltung einer photolabilen Gruppe, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“ H.
20 Giegrich, 1996, Konstanz) ab.

Nach der Spaltung verbleibt an der Base ein Linkerrest (A) (Fig.9c). Falls die nach der Spaltung am Linkerrest frei gewordene Mercapto-Gruppe weitere Reaktionen stört, kann sie mit bekannten Mitteln chemisch modifiziert werden (wie z.B. durch Disulfid-
25 oder Iodacetatverbindungen).

Insgesamt spielen die Größe, die Ladung und die chemische Struktur des Markers, die Länge des spaltbaren Linkers und des Linker-Rests sowie auch die Wahl der Polymerase eine wichtige Rolle. Sie bestimmen gemeinsam, ob das markierte NT^{*}
30 durch die Polymerase in die wachsende Nukleinsäurekette eingebaut wird, und ob

dadurch der Einbau des nächsten markierten NT^{*} verhindert wird. Zwei Bedingungen sind dabei besonders zu berücksichtigen:

Einerseits ist es wichtig, dass die Polymerase die Nukleinsäurekette mit dem
5 eingebauten modifizierten NT^{*} nach der Spaltung des Linkers weiter verlängern kann.
Es ist also wichtig, dass der Linkerrest "A" (Fig. 9c) nach der Spaltung keine
wesentliche Störung für die weitere Synthese darstellt. Andererseits müssen
eingebaute, nicht gespaltene NTs^{*} ein Hindernis darstellen. Es können viele für die
Reaktion geeignete NTs^{*} synthetisiert werden. Im einzelnen muß für jede Kombination
10 aus Polymerase und NTs^{*} eine Vorversuchsreihe durchgeführt werden, in der die
Tauglichkeit eines bestimmten NT^{*}-Typs für die Sequenzierung erprobt wird.

Die Pufferbedingungen werden nach Angaben des Polymeraseherstellers gewählt.
Die Reaktionstemperatur wird für nicht thermostabile Polymerasen nach Angaben des
15 Herstellers gewählt (z.B. 37°C für Sequenase Version 2), für thermostabile
Polymerasen (z.B. Taq-Polymerase) beträgt die Reaktionstemperatur maximal den
Temperaturwert (x). Dieser Temperaturwert (x) hängt von der T_m des Genprodukt-
Primer-Komplexes und wird z.B. als T_m (Genprodukt-Primer-Komplex) minus 5°C
errechnet (z.B. T_m ist 47°C, dann liegt die maximale Reaktionstemperatur bei 42°C).
20 Diese Pufferbedingungen und Reaktionstemperatur werden weiter als "optimale
Puffer- und Temperaturbedingungen" bezeichnet.

Die Reaktionszeit (entspricht der Dauer des Einbau-Schrittes in einem Zyklus) beträgt
vorzugsweise weniger als eine Stunde, idealerweise liegt die Reaktionszeit zwischen
25 10 sec und 10 min.

Als Beispiele von geeigneten Kombinationen zwischen NT^{*} und Polymerase sind
folgende Kombinationen zu nennen:

30 Falls DNA (z.B. cDNA) in die Reaktion eingesetzt wird, können NT^{*} mit einem kurzen
Linkerrest (Fig. 9e,h,i): dNTP-SS-TRITC (L7), dNTP-SS-Cy3 (L11) und / oder NT^{*} mit

einem langen Linkerrest (Fig. 9f,g,j): dNTP-SS-TRITC (L14) in Kombination mit Sequenase Version 2, Taq-Polymerase (GibcoBRL), ProHA-DNA-Polymerase (Eurogentec) oder Klenow Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech) eingesetzt werden.

5

Falls RNA (z.B. mRNA) in die Reaktion eingesetzt wird, können NT* mit einem kurzen Linkerrest (Fig. 9e,h,i): dNTP-SS-TRITC (L7), dNTP-SS-Cy3 (L11) und / oder NT* mit einem langen Linkerrest (Fig. 9f,g,j): dNTP-SS-TRITC (L14) in Kombination mit AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse
10 Transcriptase ohne RNase-Aktivität eingesetzt werden.

Synthesen:

Modifiziertes dUTP mit einem langen spaltbaren Linker (Fig. 9f-1) Als Ausgangssubstanzen dienen 5-(3-Aminoallyl)-2'-deoxyuridin- 5'-triphosphat, AA-
15 dUTP, (Sigma), 3,3'-Dithio-bis(propionsäure- N-Hydroxysuccinimidester), DTBP-NHS, (Sigma), 2-Mercaptoethylamin, MEA, (Sigma). Zu 100 µl 50mmol/l Lösung von AA-dUTP in 100mmol/l Borat-Puffer pH 8.5 werden 3 Äquivalente an DTBP-NHS in DMF (25 µl 0.4mol/l Lösung) zugegeben. Das Reaktionsgemisch wird 4 Stunden bei RT. inkubiert. Anschließend wird konz. Ammoniumacetat-Lösung (pH 9) zugegeben bis
20 die Gesamtkonzentration an CH₃COONH₄ in der Reaktionslösung 100mmol/l ist, und die Reaktion wird eine weitere Stunde inkubiert. Danach werden zu diesem Gemisch 200 µl 1mol/l MEA-Lösung, pH 9, zugegeben und eine Stunde bei RT inkubiert. Anschließend wird zu diesem Gemisch solange eine gesättigte Lösung an I₂ in 0.3M KI-Lösung zugetropft, bis die Jodfarbe bestehen bleibt. Die modifizierten Nukleotide
25 werden auf einer DEAE-Cellulose-Säule in Ammoniumcarbonat-Gradient (pH 8.5) von anderen Reaktionsprodukten abgetrennt. Isolierung des Nukleotids mit dem spaltbaren Linker erfolgt auf RP-HPLC. An diesen Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Waggoner
30 Method in Enzymology 1995 v.246, S.362, Jameson et al. Method in Enzymology 1997, v.278, S.363).

Auch andere Nukleotidanaloga (z.B. nach Hobbs et al. US Patent 5,047,519, Khan et al. US Patent 5,821,356) können in die Reaktion eingesetzt werden, so dass Nukleotidanaloga mit Strukturen in Fig. 9f-2,3,4 und 9g-1,2 erzeugt werden können.

5

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung von TRITC (Tetramethylrhodamin-5-isothiocyanat, Molecular Probes) angegeben (NT*-Struktur Fig. 9j):

- 10 Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 30 µl 100mmol/l Natrium-Borat-Puffer, pH 9, aufgelöst (10mmol/l NT*). Dazu werden 10 µl 10mmol/l TRITC in DMF gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT* erfolgt über RP-HPLC in Methanol-Wasser Gradient. In ähnlicher Weise können andere Farbstoffe an die Aminogruppe des Linkers gekoppelt werden.

15

Das so hergestellte NT* erfüllt die Anforderungen des Einbaus in den DNA-Strang, des Fluoreszenznachweises und Kettenabbruchs nach dem Einbau und der Aufhebung der Hemmung, die für das Gelingen des Verfahrens notwendig sind.

- 20 Beispiel der Spaltung von Disulfidverbindung im modifizierten NT*. Die Spaltung erfolgt durch eine Zugabe von 20 bis 50mmol/l DTT oder Mercaptoethanol (Sigma) Lösung pH 8 auf die Reaktionsoberfläche. Die Oberfläche wird 10 min. mit dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

25

Modifiziertes dUTP (dUTP-SS-CH₂CH₂NH₂) mit einem kurzen spaltbaren Linker (Fig. 9e-1). Als Ausgangssubstanzen dienen: Bis-dUTP, synthetisiert nach Hanna (Method in Enzymology 1989, v.180, S.383), 2-Mercaptoethylamine, MEA, (Sigma).

- 30 Zu 400 µl 100mmol/l Bis-dUTP in 40mmol/l Boratpuffer pH 8.5 werden 100µl 100mmol/l MEA-Lösung, pH 8.5, in H₂O zugegeben und 1 Stunde bei RT inkubiert.

- 60 -

Anschließend wird zu diesem Gemisch solange eine gesättigte Lösung an I_2 in 0.3mol/l KI-Lösung zugetropft, bis die Jodfarbe bestehen bleibt. Die Nukleotide (Bis-dUTP und dUTP-SS-CH₂CH₂NH₂) können z.B. durch eine Ethanol-Präzipitation oder auf einer DEAE-Cellulose-Säule in Ammoniumcarbonat-Gradient (pH 8.5) von
 5 anderen Reaktionsprodukten abgetrennt. Bis-dUTP stört bei der anschließenden Ankopplung eines Farbstoffs an die Aminogruppe des Linkers nicht, so dass die Abtrennung des dUTP-SS-CH₂CH₂NH₂ von bis-dUTP im Endreinigungsschritt erfolgen kann.

- 10 In einer ähnlichen Weise kann auch dCTP (Fig.9-e2) modifiziert werden, dabei dient Bis-dCTP als Ausgangssubstanz (synthetisiert nach Hanna et al. Nucleic Acid Research 1993, v.21, S.2073).

Weitere NT* (dUTP* und dCTP*) mit einem kurzen Linkerrest können in einer
 15 ähnlichen Weise synthetisiert werden, wobei NT* beispielsweise folgende Strukturen aufweisen können (Fig.9e):

dUTP-SS-(CH₂)_n-NH₂, Fig.9e-1,

dCTP-SS-(CH₂)_n-NH₂, Fig.9e-2,

wo n zwischen 2 und 6 liegt, vorzugsweise zwischen 2 und 4, weitere Beispiele sind:

20 dUTP-SS-(CH₂)_n-X-CO-(CH₂)_m-Z,

dUTP-SS-(CH₂)_n-X-CO-Y-(CH₂)_m-Z,

dCTP-SS-(CH₂)_n-X-CO-(CH₂)_m-Z,

dCTP-SS-(CH₂)_n-X-CO-Y-(CH₂)_m-Z,

X = NH, O, S

25 Y = NH, O, S

Z = NH₂, OH, Farbstoff

wo (n + m) zwischen 4 und 10 liegt, vorzugsweise zwischen 4 und 6.

An den Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden
 30 ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996,

R.Haugland, Molecular Probes, Waggoner Method in Enzymology 1995 v.246, S.362, Jameson et al. Method in Enzymology 1997, v.278, S.363).

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung des
5 FluoroLink™ Cy3 monofunktional dye (Amersham Pharmacia biotech) (NT*-Struktur
Fig. 9i) angegeben. Das ist ein monofunktionaler NHS-Ester-Fluoreszenzfarbstoff. Die
Reaktion wird nach Angaben des Herstellers durchgeführt:

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 300 µl 100mmol/l
Natrium-Borat-Puffer pH 8.5 aufgelöst. Dazu wird Farbstoff (300nmol) gegeben und
10 1h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT* erfolgt
über RP-HPLC in einem Methanol-Wasser Gradienten.

Ein weiteres Beispiel der Ankopplung eines Farbstoffs an den Linker wird die
Ankopplung von TRITC (Tetramethylrhodamin-5-isothiocyanat, Molecular Probes)
15 angegeben (dUTP-SS-TRITC Fig.9h).

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 30 µl 100mmol/l
Natrium-Borat-Puffer pH 9 aufgelöst (10mmol/l NT*). Dazu werden 10 µl 10mmol/l
TRITC in DMF gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff
20 modifizierten NT* erfolgt über RP-HPLC in einem Methanol-Wasser Gradienten.

Das so hergestellte NT* erfüllt die Anforderungen des Einbaus in den DNA-Strang,
des Fluoreszenznachweises und Kettenabbruchs nach dem Einbau und der
Aufhebung der Hemmung, die für das Gelingen des Verfahrens notwendig sind.

25

Beispiel der Spaltung der Disulfidverbindung im modifizierten NT*. Die Spaltung erfolgt
durch Zugabe von 20 bis 50mmol/l Dithiothreitol-Lösung (DTT) oder Mercaptoethanol-
Lösung (Sigma), pH 8, auf die Reaktionsoberfläche. Die Oberfläche wird 10 min. mit
dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer
30 Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

Allgemeine NT-Struktur mit einer an die Ribose bzw. 2'-Deoxyribose gekoppelten zur Termination führenden Gruppe:

- In den Verfahren können unterschiedliche NT*s verwendet werden (vorzugsweise
- 5 2'-deoxy-Nukleotid-Triphosphate), die an ihrer 3'-Position des Riboseringes einen Substituenten (die zur Termination führende Gruppe) tragen. Dieser Substituent kann alleine oder zusammen mit dem Fluoreszenzfarbstoff zur Termination der Einbaureaktion führen und kann unter milden Bedingungen vom Nukleotid abgespalten werden. An diesen Substituenten ist ein für das jeweilige NT*
- 10 charakteristischer Fluoreszenzfarbstoff angekoppelt, so dass der Substituent auch die Rolle eines Linkers zwischen dem Nukleotid und dem Fluoreszenzfarbstoff übernimmt. Der Fluoreszenzfarbstoff wird vorzugsweise an diesen Linker durch eine unter milden Bedingungen spaltbare Bindung angekoppelt.
- Unter „milden Bedingungen“ werden Spaltungsbedingungen verstanden, die weder
- 15 zur Denaturierung des Primer-Nukleinsäure-Komplexes führen, noch zur Spaltung seiner einzelner Bestandteile.

Formeln (1-3) stellen Beispiele für die reversiblen spaltbaren Terminatoren dar:

- 20 1) NT-3'-O-S(1)-F
 2) NT-3'-O-S(2)-N-F
 3) NT-3'-O-S(2)-N-L-F

25 NT-3'-O - stellt den 2'-Deoxy-Nukleosid-Triphosphat-Rest dar.

S(1) - stellt einen Substituenten (Formel 1) dar, der unter milden Bedingungen vom NT* abgespalten werden kann. An diesen Substituenten ist ein Fluoreszenzfarbstoff (F) gekoppelt.

30 S(2)-N - stellt einen weiteren Substituenten (Formel 2 und 3) dar, der unter milden Bedingungen vom NT* abgespalten werden kann. Dieser Substituent ist mit

dem Fluoreszenzfarbstoff (F) durch eine unter milden Bedingungen spaltbare Gruppe (N) verbunden. Der Fluoreszenzfarbstoff kann unmittelbar an die spaltbare Gruppe (Formel 2) oder durch einem weiteren Linker (L) (Formel 3) gekoppelt sein.

- 5 Beispiele für NT*-Strukturen, NT*-Synthese, zur Polymerase-Wahl für die Einbaureaktion, Reaktionsbedingungen der NT*-Einbaureaktion und Abspaltungsreaktion sind in (Kwiatkowski WO-Patent 01/25247, Kwiatkowski US-Patent 6.255.475, Conard et al. US-Patent 6.001.566, Dower (US Patent 5.547.839), Canard et al. (US Patent 5.798.210), Rasolonjatovo (Nucleosides & Nucleotides 1999, v.18 S.1021), Metzker et al. (NAR 1994, v.22, S.4259), Welch et al. (Nucleosides & Nucleotides 1999, v.18, S.197) beschrieben.

Spaltbare Bindung zwischen dem Nukleotid und dem Substituenten, Spaltung:

- Der zur Termination führende Substituent ist an das NT durch eine unter milden Bedingungen spaltbare Bindung gekoppelt.

Beispiele für diese Verbindungen stellen Ester und Acetale dar.

Die Spaltung der Ester erfolgt vorzugsweise im basischen pH-Bereich (z.B. 9 bis 11).

Die Spaltung von Acetalen erfolgt im sauren Bereich (z.B. zwischen 3 und 4).

- Ester können auch enzymatisch durch Polymerasen oder Esterasen abgespalten werden.

In einer bevorzugten Ausführungsform der Erfindung wird der Substituent zusammen mit dem Fluoreszenzfarbstoff in einem Schritt abgespalten.

- 25 Spaltbare Bindung zwischen dem Substituenten und dem Fluoreszenzfarbstoff, Spaltung:

In einer anderen bevorzugten Ausführungsform der Erfindung ist der Fluoreszenzfarbstoff an den Substituenten durch eine unter milden Bedingungen spaltbare Gruppe gekoppelt.

- 30 Vorzugsweise gehört die genannte Gruppe zu chemisch oder enzymatisch spaltbaren oder photolabilen Verbindungen.

Ester-, Thioester-, Disulfid-Verbindungen und photolabile Verbindungen eignen sich besonders gut als spaltbare Verbindung zwischen dem Substituenten und dem Fluoreszenzfarbstoff.

- 5 Als Beispiele von chemisch spaltbaren Gruppen sind Ester-, Thioester- und Disulfid-Verbindungen bevorzugt („Chemistry of protein conjugation and crosslinking“ Shan S. Wong 1993 CRC Press Inc., Herman et al. Method in Enzymology 1990 v.184 S.584, Lomant et al. J.Mol.Biol. 1976 v.104 243, "Chemistry of carboxylic acid and esters" S.Patai 1969 Interscience Publ.). Beispiele für photolabile Verbindungen
- 10 können in folgenden Literaturstellen gefunden werden: "Protective groups in organic synthesis" 1991 John Wiley & Sons, Inc., V. Pillai Synthesis 1980 S.1, V. Pillai Org.Photochem. 1987 v.9 S.225, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“ H. Giegrich, 1996, Konstanz, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“
- 15 S.M.Bühler, 1999, Konstanz).

Der Spaltungsschritt ist in jedem Zyklus vorhanden und muß unter milden Bedingungen verlaufen, so dass die Nukleinsäuren nicht beschädigt oder modifiziert werden.

20

- Die Spaltung läuft bevorzugt chemisch (z.B. in milder saurer oder basischer Umgebung für eine Ester-Verbindung oder durch Zugabe eines Reduktionsmittels, z.B. Dithiothreitol oder Mercaptoethanol (Sigma) bei der Spaltung einer Disulfid-Verbindung), oder physikalisch (z.B. durch Beleuchtung der Oberfläche mit Licht
- 25 einer bestimmten Wellenlänge für die Spaltung einer photolabilen Gruppe, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“ H. Giegrich, 1996, Konstanz) ab.

- In dieser Ausführungsform wird nach der Detektion zunächst der
- 30 Fluoreszenzfarbstoff abgespalten und erst dann der an die 3'-Position gekoppelte, zur Termination führende Substituent.

Die Erfindung soll an einigen schematischen Figuren weiter verdeutlicht werden.
Legenden zu Figuren:

Fig. 1 Eine schematische Darstellung einer Ausführungsform des
5 Sequenzierautomaten

- 101 Lichtquelle für Epifluoreszenzmodus
- 102 Fokussierungsoptik (1)
- 103 Shutter (S1)
- 10 104 Lichtstrahl des Anregungslichts
- 105 Filtersatz bzw. mehrere Filtersätze zur Selektion von Lichtwellenlängen und Farbteiler
- 106 Objektiv
- 107 Reaktionsplattform mit
 - 15 107a Pumpe
 - 107b Vorratsbehälter
 - 107c Ventile
- 108 Translationstisch (Scantisch)
- 109 Kondensor
- 20 110 Spiegel
- 111 Shutter (S2)
- 112 Fokussierungsoptik (2)
- 113 Lichtquelle für Transmissionmodus
- 114 Lichtstrahl des Transmissionslichts
- 25 115 Tubus-Optik-1
- 116 Detektionsvorrichtung
- Gehäuse ist nicht gezeigt.

Fig. 2 Flußdiagramm mit einem **Beispiel für den Ablauf wesentlicher Arbeitsschritte:**

Bei **Initialisierung** werden durch den Benutzer die Parameter für die
5 Sequenzierungsreaktion ausgewählt. Es werden folgende Parameter eingestellt:

- 1) die Art der Untersuchung, z.B. Sequenzierung langer NSKs oder Genexpressionsanalyse.
- 2) durchschnittliche Länge der immobilisierten NSKs bzw. NSKFs.
- 3) durchschnittliche Anzahl der eingebauten NT*s pro NSK.
- 10 4) Sensitivität und Spezifität der Analyse.

Im Abschnitt **präzyklische Reaktionen** werden NSKs bzw. NSKFs in MFK in Form von NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen fixiert. Das Ziel dieses Abschnittes ist die Immobilisierung der zu untersuchenden Proben in optimaler
15 Dichte (s. Beispiel Immobilisierung). Die Parameter des Hybridisierungsschrittes (Primer und PBS-Zusammensetzung, Lösungszusammensetzung, optimale Hybridisierungs- und Waschtemperatur, Primerimmobilisierungsdichte auf der Oberfläche, Konzentration der NSKs) sind vorzugsweise bekannt und bestimmen zusammen mit der Dauer des Hybridisierungsschrittes die Immobilisationsdichte der
20 NSKs.

Im Abschnitt **zyklische Reaktionen** werden markierte NT*s in den komplementären Strang immobilisierter NSKs bzw. NSKFs eingebaut und die Signale von eingebauten NT*s durch das Abscannen der Reaktionsoberfläche detektiert,
25 identifiziert und zur jeweils bestimmten NT*-Art zugeordnet (Signalverarbeitung).

Im Abschnitt **Datenverarbeitung** erfolgt die Zusammensetzung der Sequenzen aus einzelnen identifizierten und zugeordneten NT*s.

Fig. 3 stellt ein "Stand der Technik" Epi-Fluoreszenzmikroskop dar, das in den Sequenzierautomaten integriert werden kann.

117 Ableitungsoptik zum Okular

118 Tubus-Optik-2

5 119 Okular-Optik

Gehäuse ist nicht gezeigt.

Fig. 4a Eine vorteilhafte Ausführungsform der Detektionsapparatur.

Sie zeichnet sich dadurch aus, dass

10 1) mehrere Filtersätze in einem Filterrevolver oder Filterschieber 120 angebracht sind,

2) der Scantisch 108, der Filterrevolver oder der Filterschieber 120, der Shutter 103 und 111 Thermostateinheit, mit der Pumpe und den Steuerungsventilen (in Fig. nicht gezeigt) zur Steuerung der Arbeitsschritte mit dem Computer 121 verbunden

15 sind.

Zur Fokuseinstellung und Justierung wird in dieser Ausführungsform das Transmissionslicht verwendet.

Fig. 4b Diese beispielhafte Ausführungsform zeichnet sich durch folgende

20 Merkmale aus

Der Sequenzierautomat besitzt eine Vorrichtung (122) zur Intensitätskontrolle und Intensitätsregulation des Anregungslichtes. Diese Intensitätsregulierung kann beispielsweise durch Änderung der Leistung der Lichtquelle (101) erfolgen. Die
25 Vorrichtung stellt einen Teil des Regelkreises für die Lichtintensität dar und ist mit der zentralen Recheneinheit verbunden.

2) Zur Fokuseinstellung und für die Justierungsbilder wird das Fluoreszenzsignal von dem mit der Reaktionsoberfläche verbundenen Muster verwendet (s. Beispiel

30 Detektion).

Fig. 5 Ein Beispiel der Detektionsapparatur charakterisiert dadurch, daß ein oder mehrere Laser (in diesem Beispiel zwei: Laser 123 und Laser 124) als Lichtquellen dienen. Diese Laser können in das Gehäuse des Sequenzierapparates integriert oder durch Faseroptik mit dem Sequenzierautomaten verbunden werden. Für
5 zeitliche Modulation des Anregungslichtes (die Belichtungszeit liegt vorzugsweise zwischen 0.1 msec und 1 sec) wird beispielsweise eine spezielle Vorrichtung 125 verwendet

Fig. 6 Beispiele für die Reaktionsplattform:

10

Fig. 6a Eine Übersichtsdarstellung der Reaktionsplattform. Dargestellt ist eine Ausführungsform mit den vier unterschiedlich markierten NT*s, die gleichzeitig in die Einbaureaktion eingesetzt werden.

201 Befestigungsplatte

15 202 zuführender Anschluß

203 ableitender Anschluß

204a Chip mit MFK

204b MFK

205 ableitender Schlauch

20 206 Ventil für Pumpe

207 Pumpe

208 a,b,c,d zuleitende Schläuche für Reaktionslösung NT*(n)

209 zuleitender Schlauch für Waschlösung

210 zuleitender Schlauch für Probenlösung

25 211 zuleitender Schlauch für Abspaltlösung

212 a,b,c,d Ventile für Reaktionslösung NT*(n)

213 Ventil für Waschlösung

214 Ventil für Probenlösung

215 Ventil für Abspaltlösung

30 216 a,b,c,d Vorratsbehälter für Reaktionslösung NT*(n)

217 Vorratsbehälter für Waschlösung

- 218 Vorratsbehälter für Probenlösung
- 219 Vorratsbehälter für Abspaltlösung
- 220 Thermostateinheit
- 221 Öffnung für Transmissionslicht
- 5 222 Abdeckplatte
- 223 Grundplatte
- 224 Sensor

Fig. 6b Eine Übersichtsdarstellung des Chips mit dem Mikroflüssigkeitskanal (MFK).

- 10 Der Kanal kann Aufweitungen und Aufspaltungen enthalten, die zur Vergrößerung der Reaktionsoberfläche führen. Die Auswahl der jeweiligen Form des MFK hängt von der Anzahl der Objektfelder ab, die abgescannt werden müssen: bei einer großen Anzahl wird man MFK mit einer größeren Reaktionsoberfläche einsetzen.

- 15 **Fig. 6c** Eine Übersichtsdarstellung der Verteilungsvorrichtung. Dargestellt ist eine Ausführungsform mit den vier unterschiedlich markierten NT*s, die gleichzeitig in die Einbaureaktion eingesetzt werden.

Fig. 6d Eine Übersichtsdarstellung der Verteilungsvorrichtung. Dargestellt ist eine

- 20 Ausführungsform mit den vier markierten NT*s, wobei nur jeweils zwei unterschiedlich markierte NT*s gleichzeitig in die Einbaureaktion im Zyklus N eingesetzt werden. Die anderen zwei werden im Zyklus N+1 eingesetzt.

Fig. 6e Eine Übersichtsdarstellung der Verteilungsvorrichtung. Dargestellt ist eine

- 25 Ausführungsform, bei der nur ein NT* pro Zyklus eingesetzt wird, wobei alle vier NT*s die gleiche Markierung tragen.

Fig. 6f Eine Übersichtsdarstellung der Reaktionsplattform. Dargestellt ist eine

- 30 Ausführungsform, bei der ein Sensor 224 den Austausch der Lösungen z.B. optisch kontrollieren kann.

Fig. 7 Schematische Übersichtsdarstellung des Scannvorganges der Reaktionsoberfläche in einem Zyklus. Dabei werden 2D Bilder (301) von mehreren Objektfeldern (302) gemacht. Fluoreszenzsignale (303) einzelner eingebauter NT*s besitzen charakteristische Koordinaten $X(n), Y(n)$.

5

Fig. 8 Detektionsschritt in einer Ausführungsform mit 4NT*s ($NT^*_{1,2,3,4}$), die mit verschiedenen Farbstoffen markiert sind

Nach der Einstellung der X,Y-Koordinaten eines Objektfeldes wird die Fokusposition der Reaktionsoberfläche überprüft bzw. eingestellt. Die Überprüfung erfolgt beispielsweise im Transmissionslichtmodus, der Shutter S2 (111) ist offen, der Shutter S1 (103) ist geschlossen. Danach erfolgt die Aufnahme der Fluoreszenzsignale. In diesem Beispiel wird für jeden Farbstoff ein spezifischer Filtersatz ($NT^*_{(n)}$) verwendet. Während der Belichtungszeit ist der Shutter S1 (103) offen und der Shutter S2 (111) geschlossen.

10
15

Fig. 9 Beispiele für Nukleotid-Strukturen, die im Verfahren eingesetzt werden.

20

25

30

Ansprüche:**Anspruch 1**

Sequenzierautomat, zur parallelen Sequenzierung einer Population einzelner auf
5 einer planen Oberfläche fixierter Nukleinsäurekettenmoleküle, wobei diese
Sequenzierung durch den sequentiellen Aufbau eines zur jeweiligen fixierten
Nukleinsäurekette komplementären Stranges mit den mit Fluoreszenzfarbstoffen
reversibel markierten Nukleotiden stattfindet, wobei dieser sequentielle Aufbau in
zyklische Reaktionen abläuft. Dieser Sequenzierautomat schließt folgende
10 Elemente ein:

- Ein optisches System zur Detektion von Signalen von einzelnen Molekülen, das
folgende Komponenten einschließt:

15 Eine Quelle der elektromagnetischen Strahlung zur Anregung der
Fluoreszenz von an die modifizierten Nukleotide gekoppelten Farbstoffen,

20 Eine Vorrichtung zur Fokussierung der zur Anregung der Fluoreszenz
eingesetzten elektromagnetischen Strahlung und zum Sammeln emittierter
elektromagnetischer Strahlung (Fluoreszenzsignale) von einzelnen
Farbstoffmolekülen, die an die modifizierten in die den zu sequenzierenden
Nukleinsäureketten komplementären Stränge eingebauten
Nukleotidmolekülen gekoppelt sind,

25 Eine Filtervorrichtung zur Selektion von Wellenlängen der zur Anregung der
Fluoreszenz eingesetzten elektromagnetischen Strahlung und gesammelter
elektromagnetischer Strahlung (Fluoreszenzsignale),

30 - Eine Detektionsvorrichtung zur Detektion der durch die Filtervorrichtung
selektierter elektromagnetischer Strahlung (Fluoreszenzsignale) von einzelnen
Farbstoffmolekülen, die an die modifizierten in die den zu sequenzierenden

Nukleinsäureketten komplementären Stränge eingebauten Nukleotidmolekülen gekoppelt sind,

- Eine Translationsvorrichtung zur Translation der Reaktionsplattformen beim
5 Abscannen der Oberfläche und zum Wechsel zwischen Reaktionsplattformen während der Zyklenschritte,
- Eine oder mehrere Reaktionsplattformen auf der Translationssvorrichtung zur Durchführung sequentieller Reaktionszyklen mit immobilisierten Nukleinsäureketten,
10 wobei diese Plattformen eine gleichzeitige Detektion der Signale von vielen einzelnen Farbstoffmolekülen erlaubt, die an die modifizierten in die den zu sequenzierenden Nukleinsäureketten komplementären Stränge eingebauten Nukleotidmolekülen gekoppelt sind,
- 15 - Ein Gehäuse zur Halterung von optischem System, Detektionsvorrichtung und Translationsvorrichtung
- Eine Analysevorrichtung zur Bestimmung von Sequenzen fixierter Nukleinsäureketten anhand durch die Detektionsvorrichtung detektierter Signale
20 von einzelnen, modifizierten in die den zu sequenzierenden Nukleinsäureketten komplementären Stränge eingebauten Nukleotidmolekülen
- Eine Steuerungsvorrichtung zur Steuerung:
 - a) der Zyklen in der Reaktionsplattform
 - 25 b) des optischen Systems
 - c) der Translationsvorrichtung
 - d) der Analysevorrichtung

Anspruch 2

Sequenzierungsautomat nach Anspruch 1 charakterisiert dadurch, dass die zur Anregung der Fluoreszenz eingesetzte elektromagnetische Strahlung im Epifluoreszenzmodus auf die Reaktionsoberfläche geleitet wird.

5

Anspruch 3

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass das optische System, die Translationsvorrichtung und das Gehäuse Teile eines Fluoreszenzmikroskopes sind.

10

Anspruch 4

Sequenzierungsautomat nach Anspruch 1 bis 3 dadurch charakterisiert, dass die Quelle der elektromagnetischen Strahlung eine Lampe ist

15 Anspruch 5

Sequenzierungsautomat nach Anspruch 4 dadurch charakterisiert, dass die Quelle der elektromagnetischen Strahlung eine Quecksilberdampf-Lampe ist

Anspruch 6

20 Sequenzierungsautomat nach Anspruch 1 bis 3 dadurch charakterisiert, dass die Quelle der elektromagnetischen Strahlung ein oder mehrere Laser sind

Anspruch 7

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass
25 Nukleinsäureketten auf einer planen Oberfläche in Form von Nukleinsäureketten-Primer-Komplexen fixiert werden

Anspruch 8

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass mehrere
30 Fluoreszenzsignale von einzelnen in verschiedenen NSKs bzw. NSKF's eingebauten NT's gleichzeitig detektiert werden.

Anspruch 9

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass mehrere Nukleinsäureketten gleichzeitig sequenziert werden.

5

Anspruch 10

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass er folgendes Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs, oder deren Fragmente, NSKFs) ausführt, bei dem man
10 eine zyklische Aufbaureaktion des komplementären Stranges der NSKs bzw. NSKFs unter Verwendung eines oder mehrerer Primer und einer oder mehrerer Polymerasen durchführt, indem man

a) zu den auf der Oberfläche gebundenen NSK-Primer-Komplexen bzw. NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere
15 Polymerasen und ein bis vier modifizierte Nukleotide (NTs*) enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs* jeweils an den NTs* befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs*
20 durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs* strukturell so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT* in denselben Strang einzubauen, man

25

b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT* verlängert werden, man

- 75 -

- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs* geeignet sind, man
- 5 d) die einzelnen, in komplementäre Stränge eingebauten NTs* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- 10 e) zur Erzeugung unmarkierter (NTs oder) NSKs bzw. NSKF's die Fluoreszenzfarbstoffe und die zur Termination führende Gruppe von den am komplementären Strang angefügten NTs* abspaltet, man
- 15 f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Gruppe geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

- 20 wobei man die relative Position einzelner NSK-Primer-Komplexe bzw. NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKs bzw. NSKF's durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

25

Anspruch 11

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass er folgendes Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs) ausführt, bei dem man

5

Fragmente (NSKFs) einzelsträngiger NSKs mit einer Länge von etwa 50 bis 1000 Nukleotiden erzeugt, die überlappende Teilsequenzen einer Gesamtsequenz darstellen können, man

10

die NSKFs unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von NSKF-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

15

eine zyklische Aufbaureaktion des komplementären Stranges der NSKFs unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

20

a) zu den auf der Oberfläche gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs*) enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs* jeweils an den NTs* befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs* durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs* strukturell so modifiziert sind, dass die Polymerase nach

25

Einbau eines solchen NT* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT* in denselben Strang einzubauen, man

30

b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT* verlängert werden, man

- 77 -

- 5 c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs* geeignet sind, man
- 10 d) die einzelnen, in komplementäre Stränge eingebauten NTs* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- 15 e) zur Erzeugung unmarkierter (NTs oder) NSKFs die Fluoreszenzfarbstoffe und die zur Termination führende Gruppe von den am komplementären Strang angefügten NTs* abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Gruppe geeignet sind, man
- 20 die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,
- wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

25

30

Anspruch 12

Sequenzierungsautomat nach Anspruch 1 dadurch charakterisiert, dass er folgendes Verfahren zur hoch parallelen Analyse der Genexpression ausführt, bei dem man

5

einzelsträngige Genprodukte bereitstellt, man

10

die Genprodukte unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von Genprodukt-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der Genprodukte unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

15

a) zu den auf der Oberfläche gebundenen Genprodukt-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs*) enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs* jeweils an den NTs* befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs* durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs* strukturell so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT* in denselben Strang einzubauen, man

20

25

b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT* verlängert werden, man

30

- 79 -

- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs* geeignet sind, man
- 5 d) die einzelnen, in komplementäre Stränge eingebauten NTs* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- 10 e) zur Erzeugung unmarkierter, (NTs oder) Genprodukte die Fluoreszenzfarbstoffe und die zur Termination führende Gruppe von den am komplementären Strang angefügten NTs* abspaltet, man
- 15 f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Gruppe geeignet sind, man
- die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,
- 20 wobei man die relative Position einzelner Genprodukt-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser Genprodukte durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt und man aus den ermittelten Teilsequenzen die Identität der Genprodukte bestimmt.

25

30

Anspruch 13

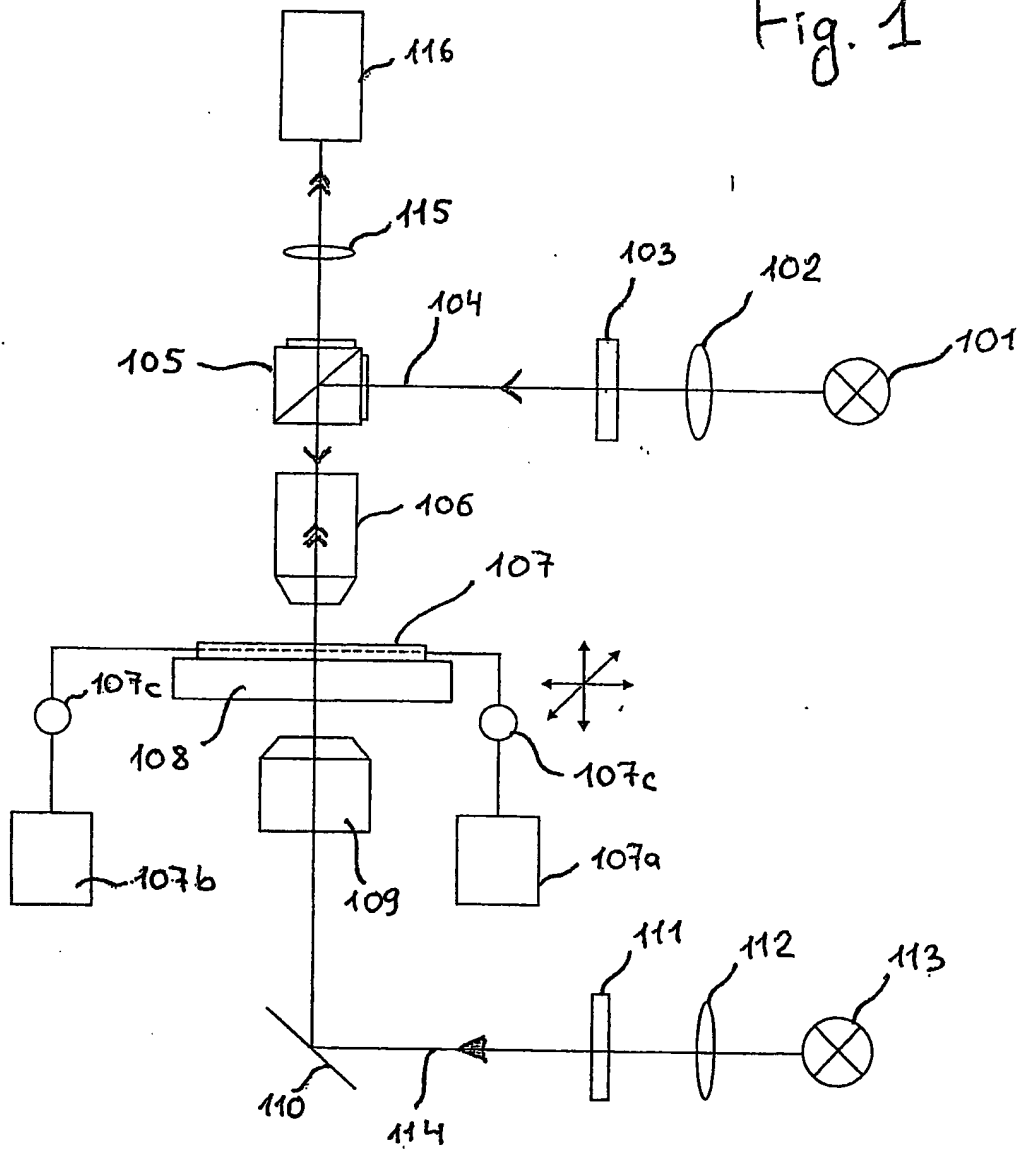
Reaktionsplattform nach Anspruch 1, zur Durchführung von Reaktionsschritten, die folgende Elemente einschließt:

- 5 - Einen austauschbaren Chip mit einem oder mehreren Mikroflüssigkeitskanälen
- Eine Verteilungsvorrichtung zur Steuerung von Lösungsaustausches im Chip
- Eine Thermostateinheit zur Steuerung der Temperatur im Chip

10 Anspruch 14

Sequenzierungsautomat nach Anspruch 1 bis 3 dadurch charakterisiert, dass die Quelle der elektromagnetischen Strahlung ein oder mehrere Laser-Dioden sind

Fig. 1



2/21

Allgemeiner Ablauf der Sequenzierungsreaktion

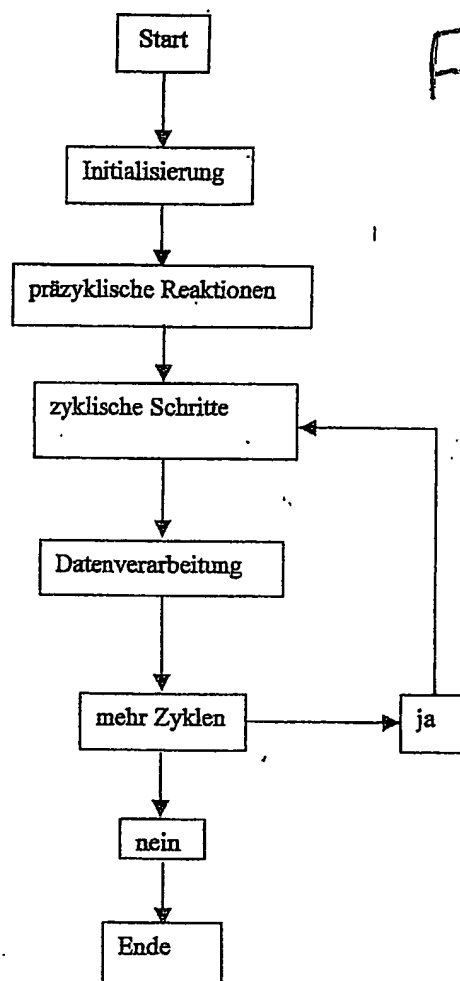


Fig. 3

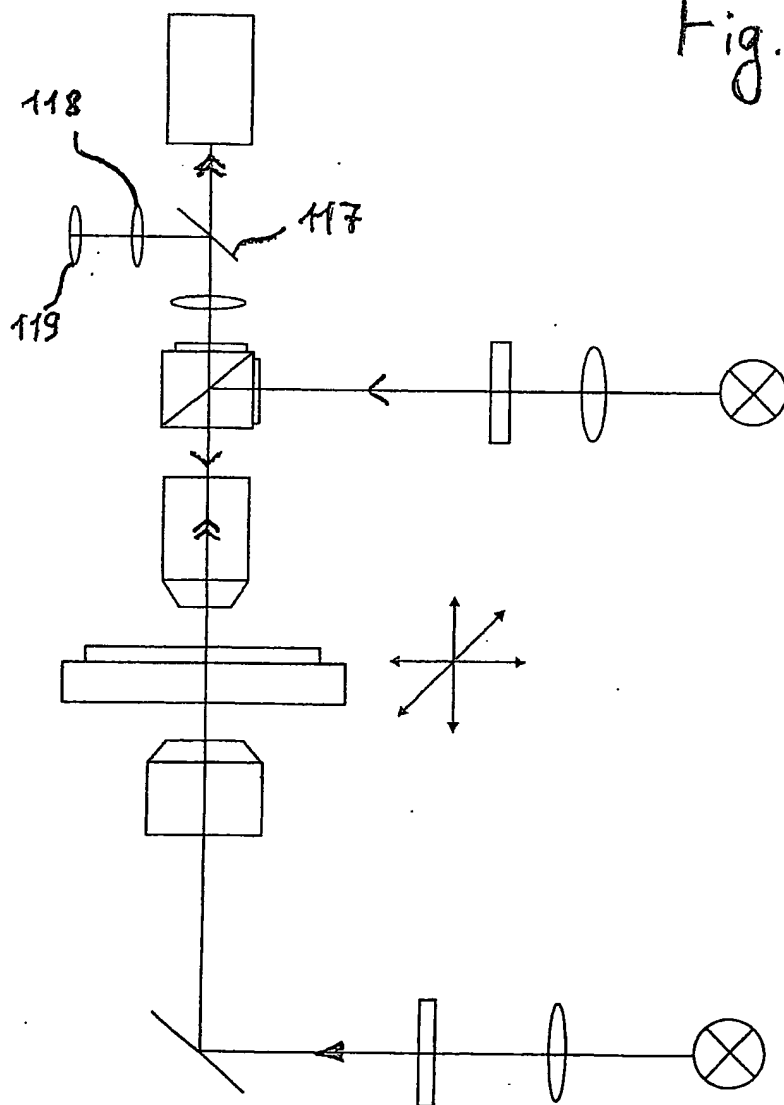


Fig. 4a

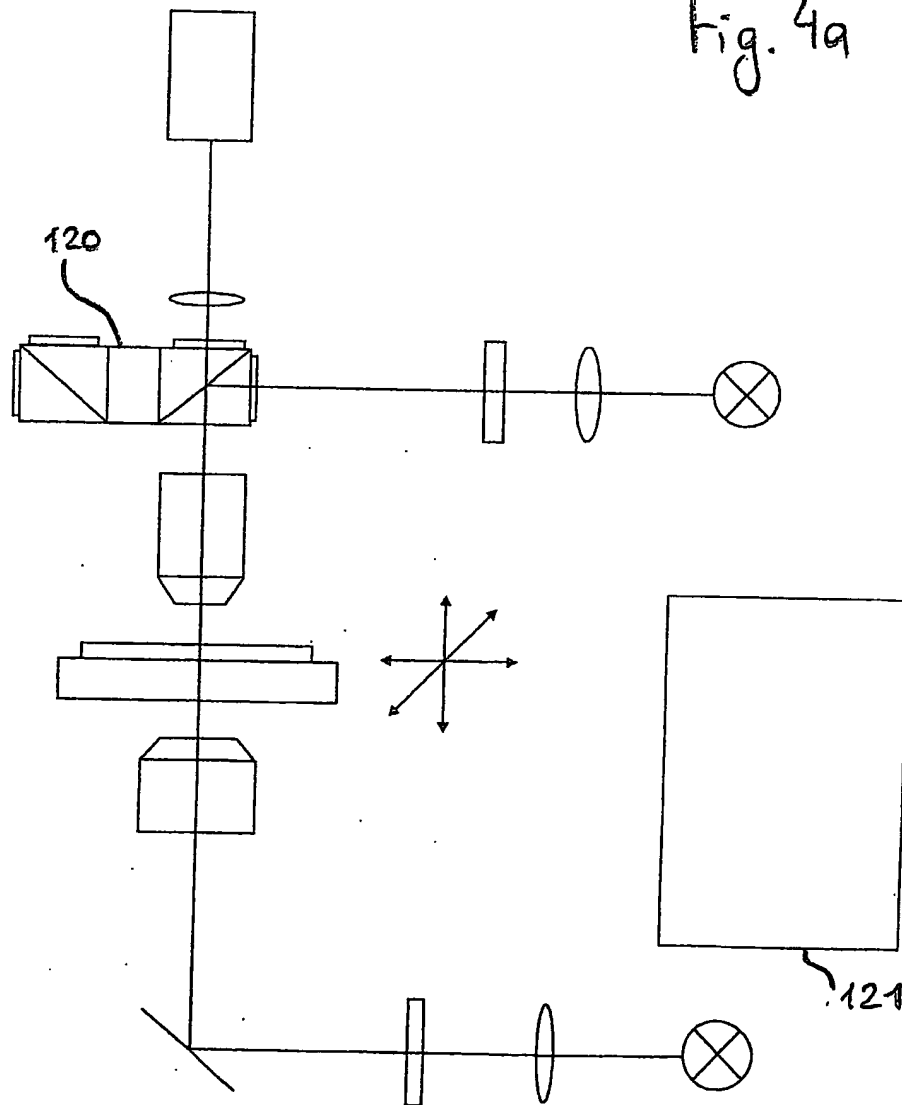
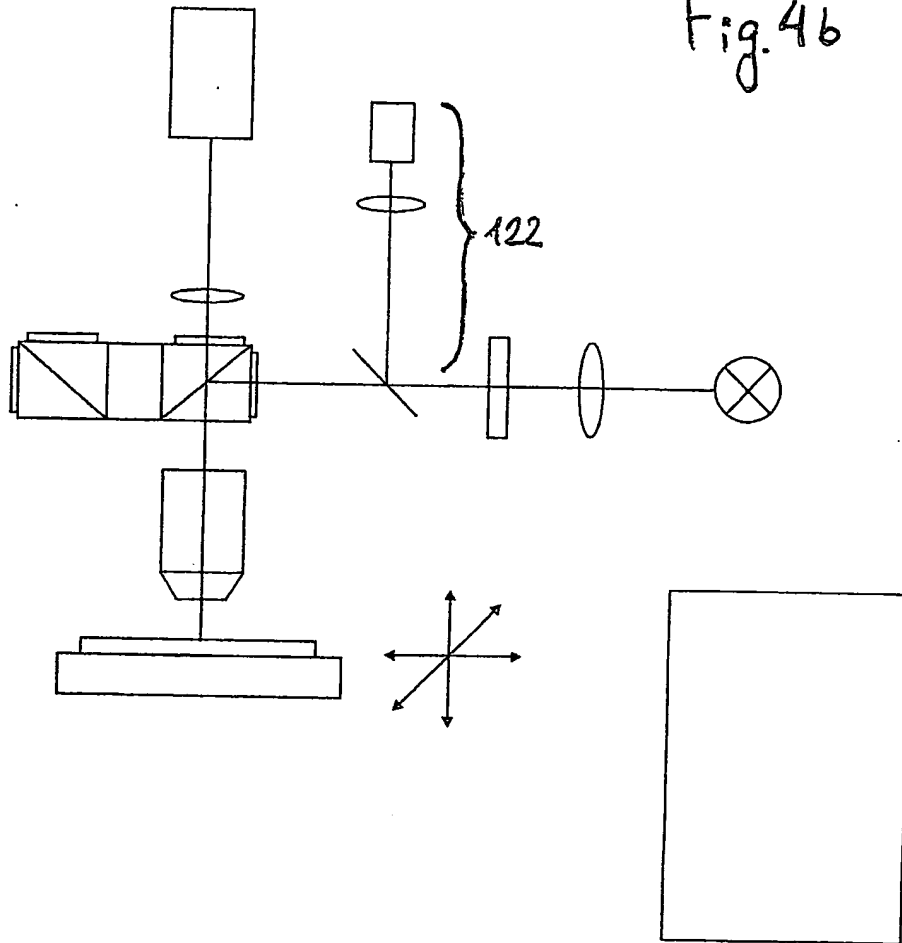
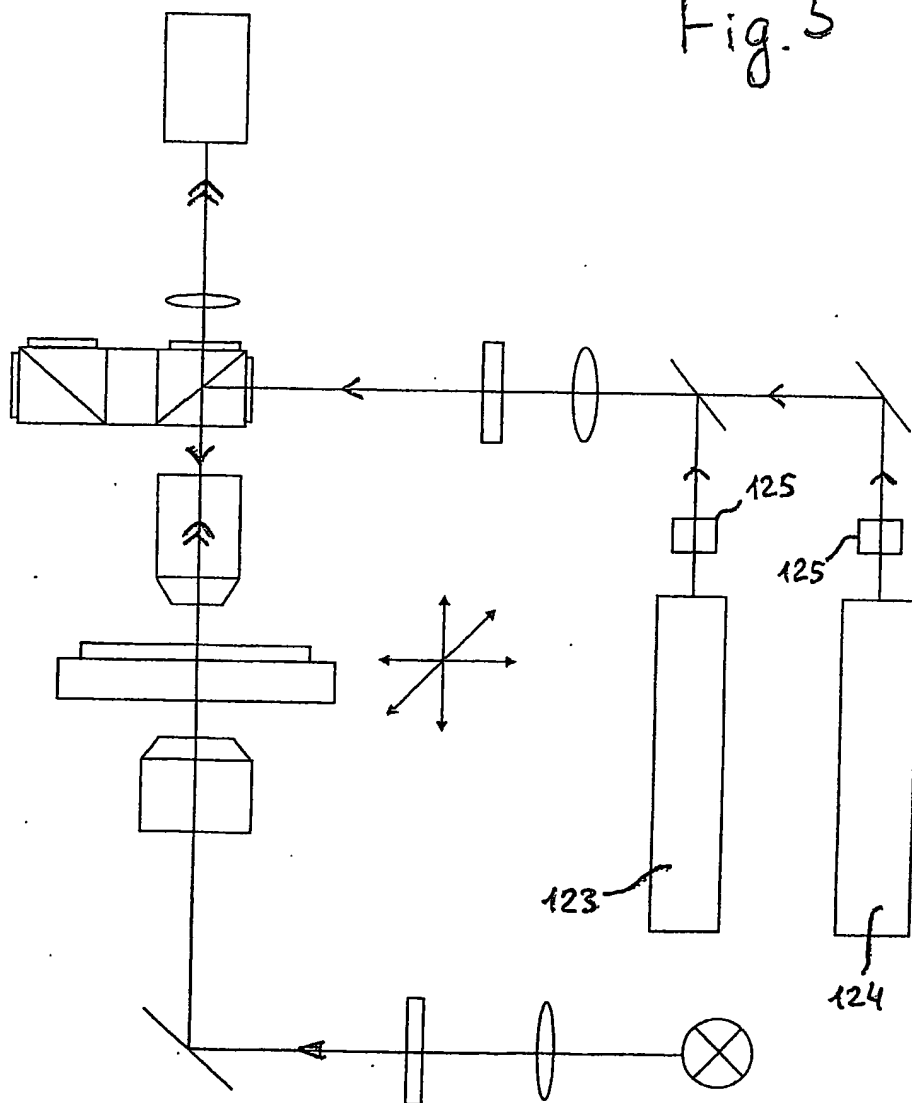


Fig. 4b

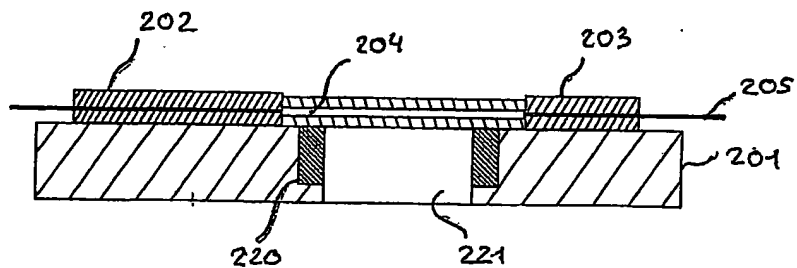
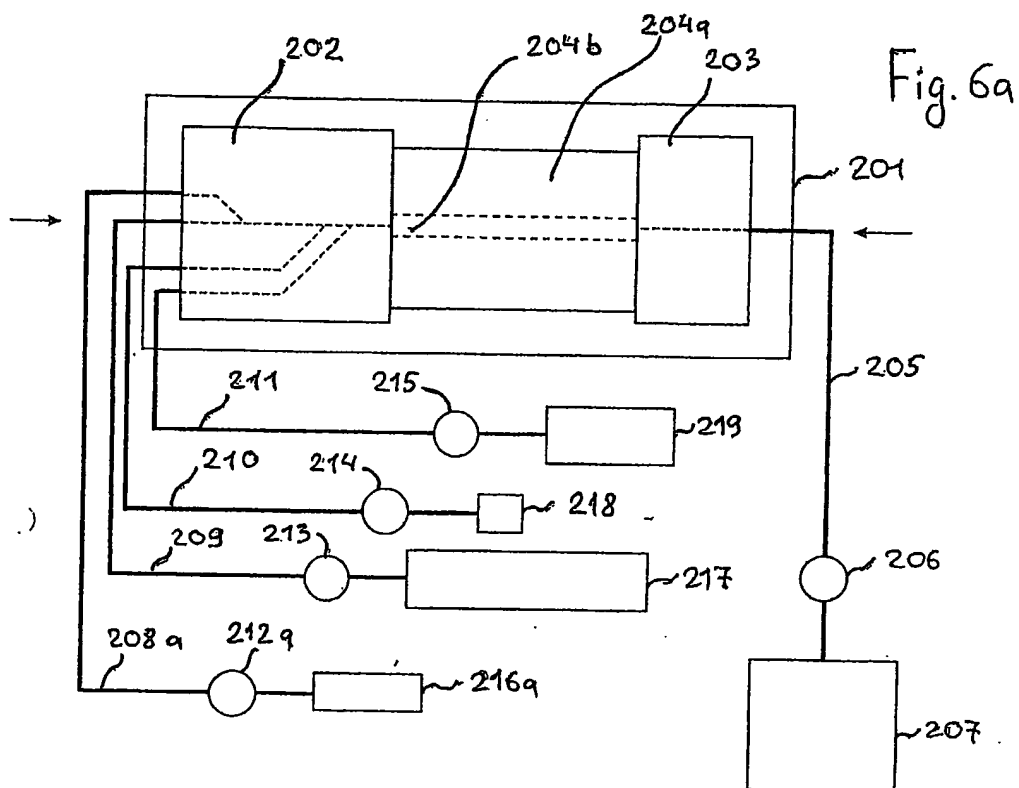


6/21

Fig. 5

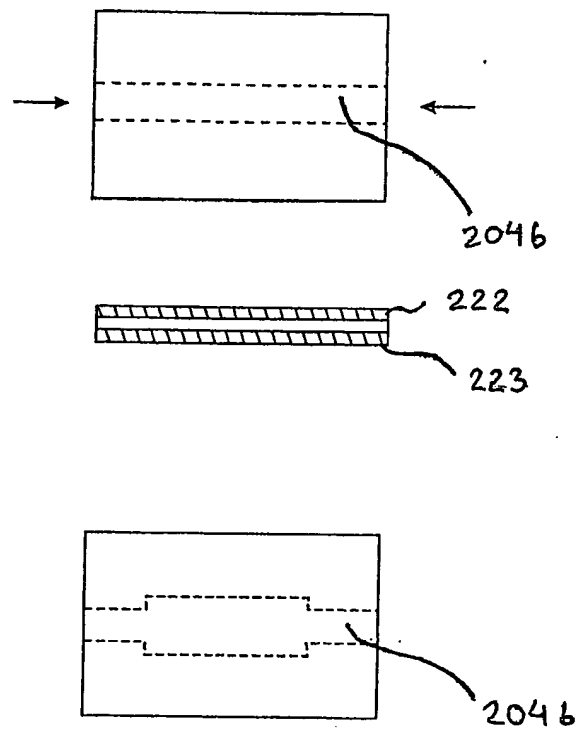


7/21

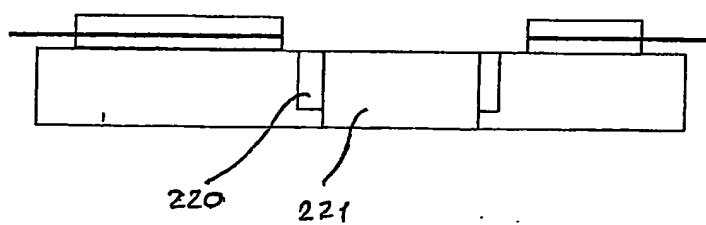
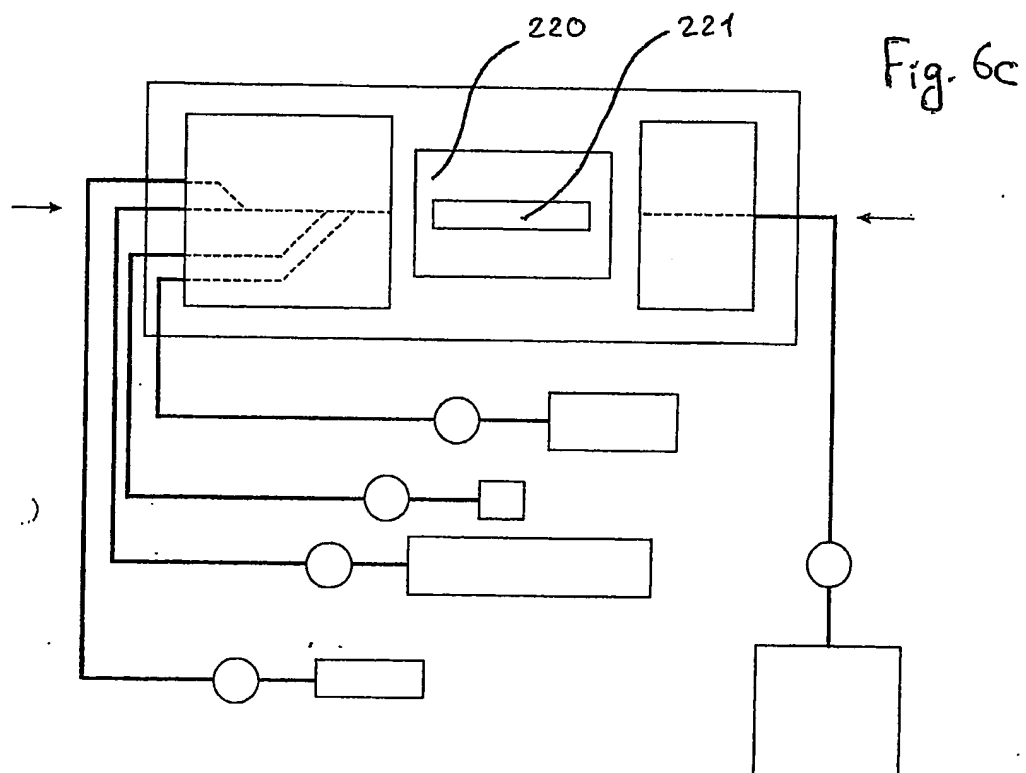


8/21

Fig. 6b

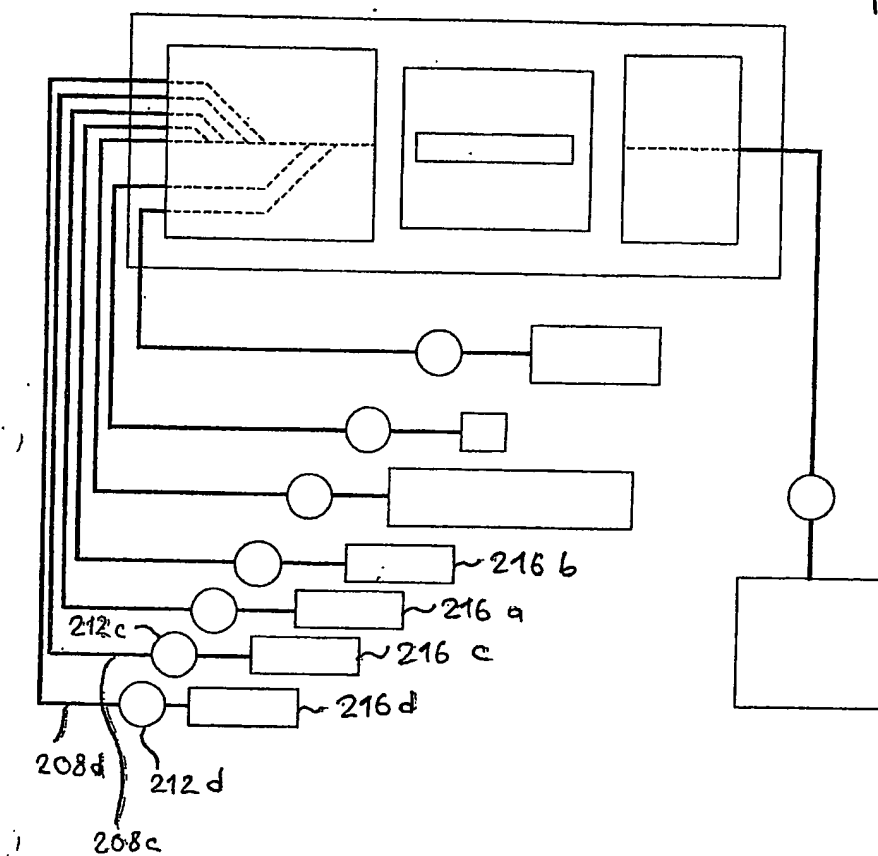


9/21



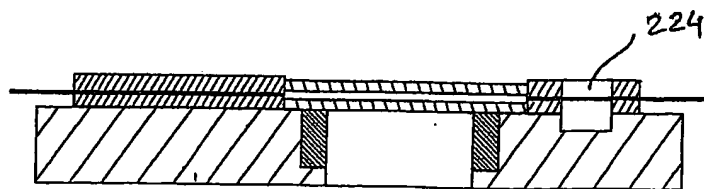
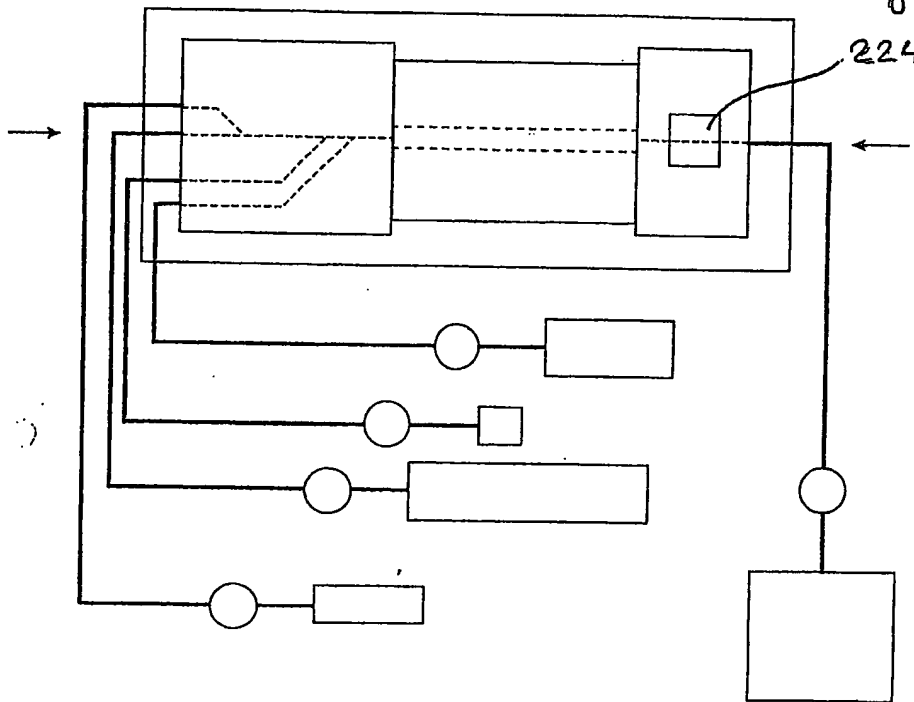
11/21

Fig. 6e



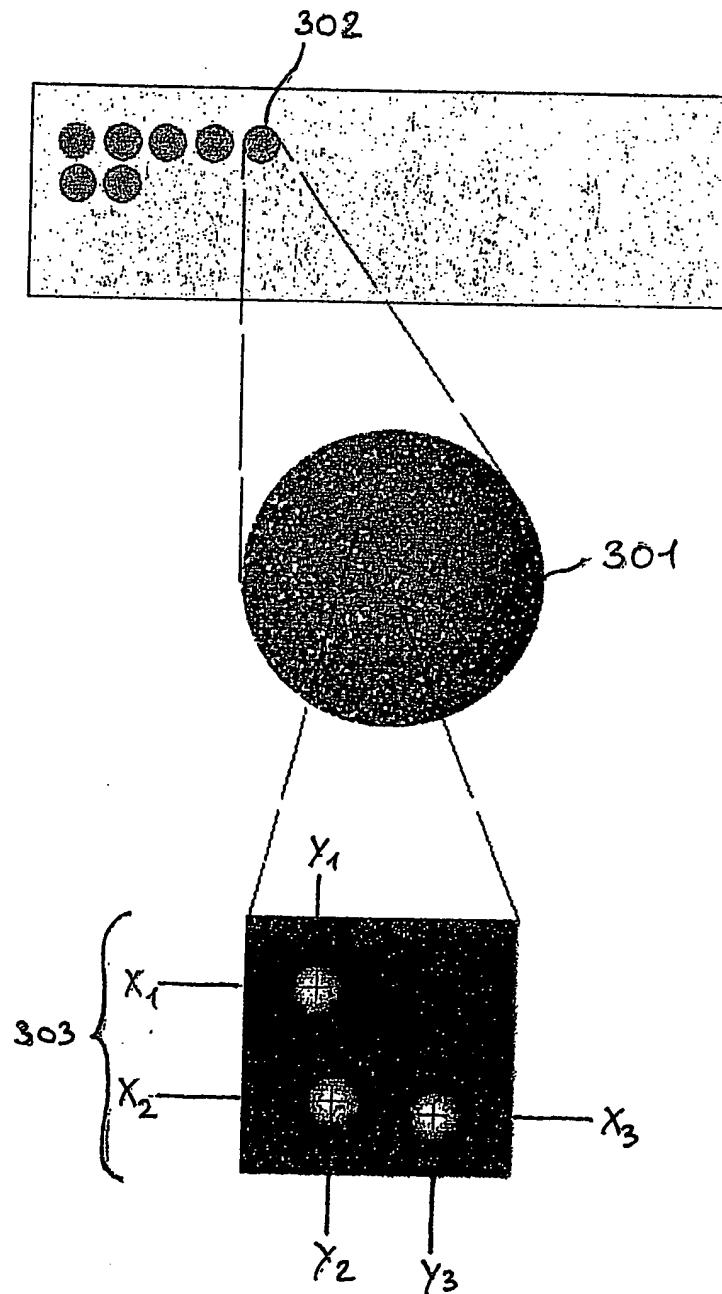
12/21

Fig. 6f



13/21

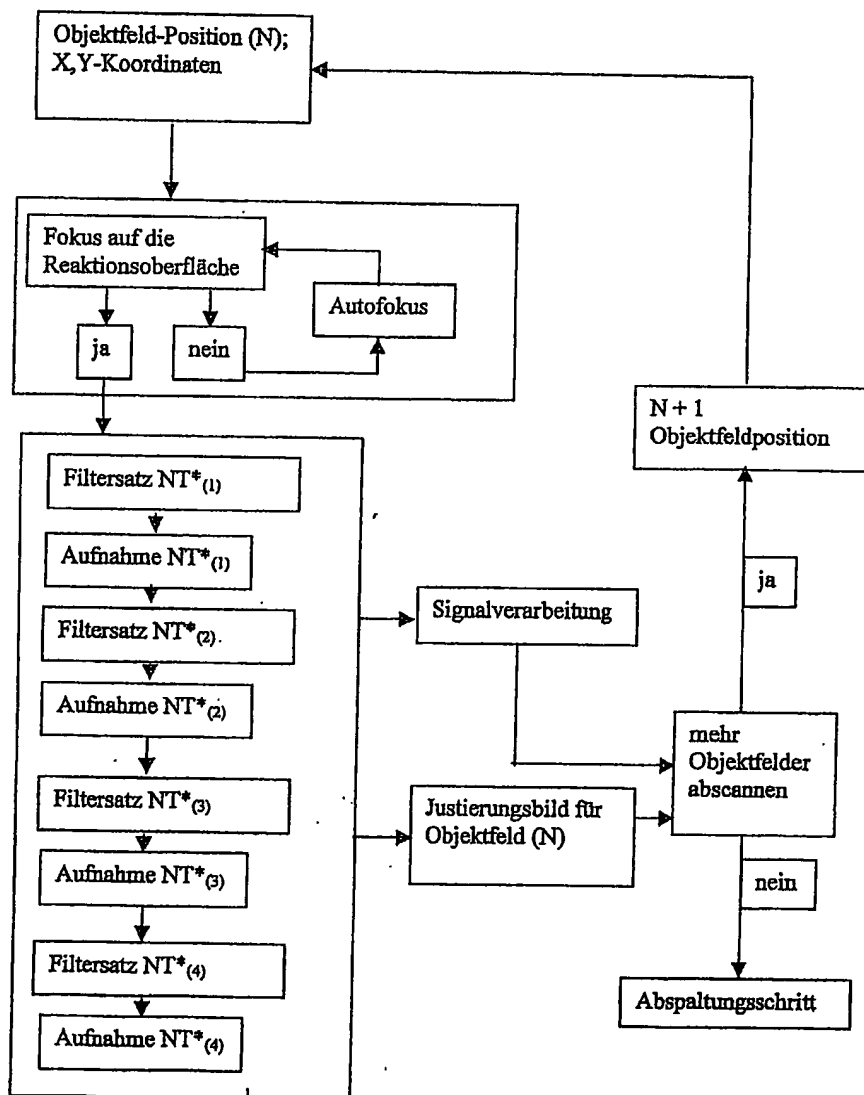
Fig. 7



14/21

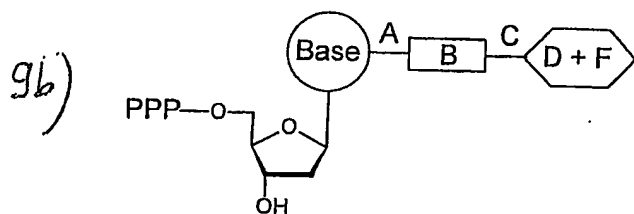
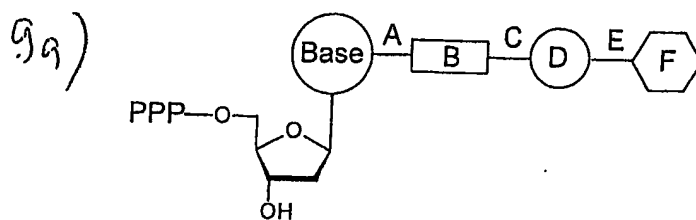
Fig. 8

Detektionsschritt in einer Ausführungsform mit 4NT*s (NT*_{1,2,3,4}), die mit verschiedenen Farbstoffen markiert sind



15/21

Fig. 9

**Legende:**

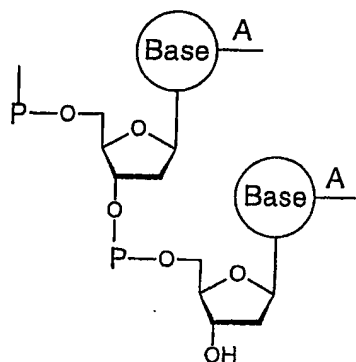
- A, C, E – Verbindungselemente im Linker
- A – Linkerrest nach der Spaltung
- B – spaltbare Verbindung / Gruppe
- D – zur Termination führende Gruppe
- F – Fluoreszenzfarbstoff (Marker)

16/21

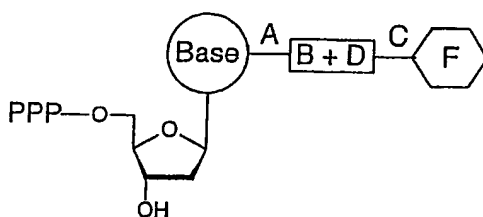
eingebaute modifizierte NTs* nach der Spaltung

Fig. 9

9c)



9d)

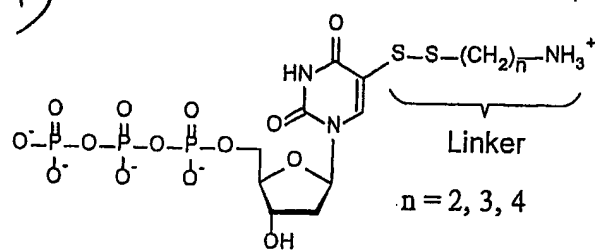
**Legende:**

- A, C, E – Verbindungselemente im Linker
- A – Linkerrest nach der Spaltung
- B – spaltbare Verbindung / Gruppe
- D – zur Termination führende Gruppe
- F – Fluoreszenzfarbstoff (Marker)

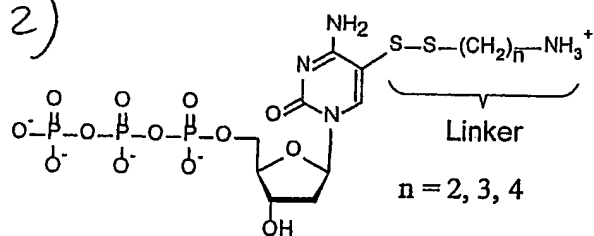
17/21

Fig. 9

ge-1)



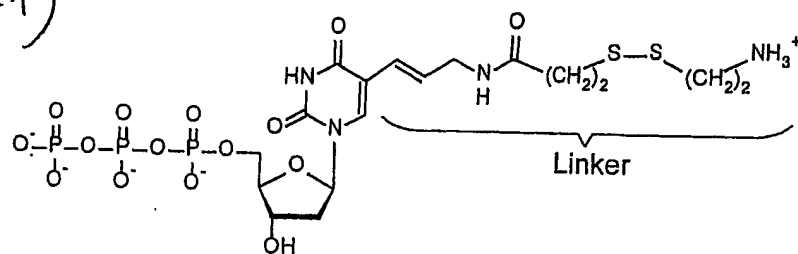
ge-2)



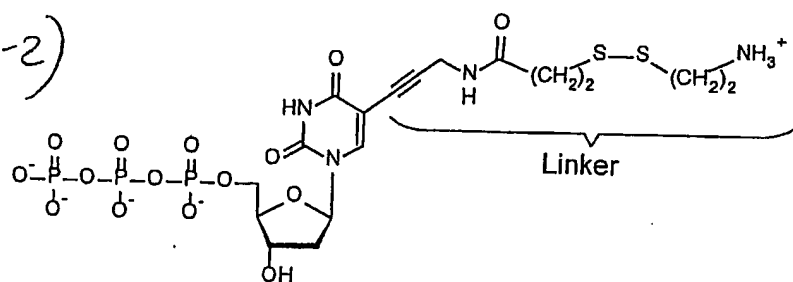
18/21

Fig. 9

9f-1)



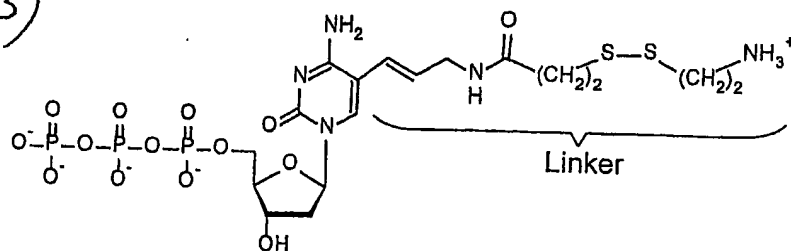
9f-2)



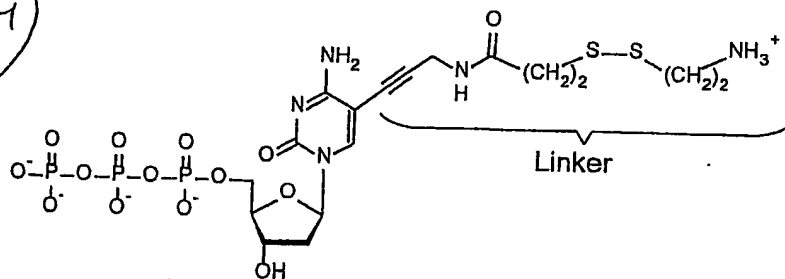
19/21

Fig. 9.

gf-3)



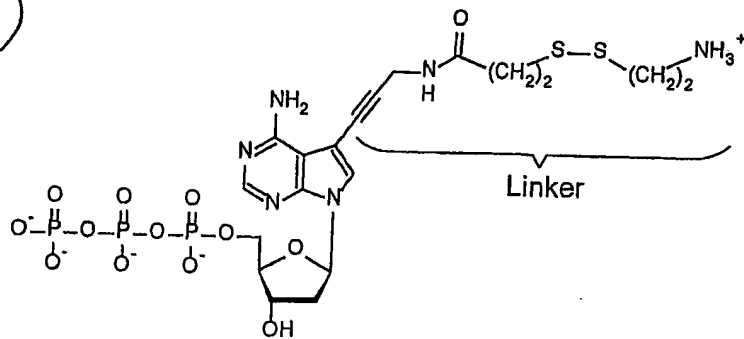
gf-4)



20/21

Fig. 9

gg-1)



gg-2)

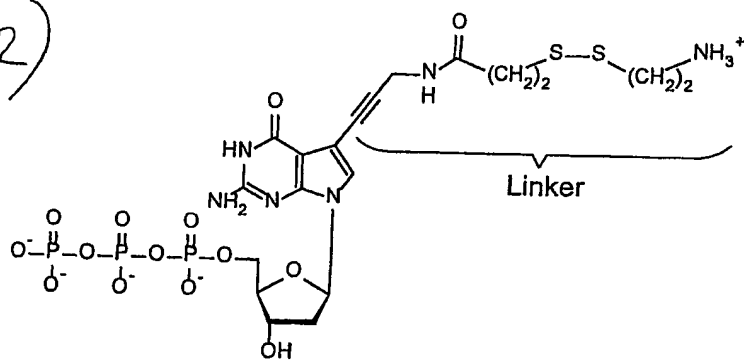


Fig. 9

